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

- Needs skills in: mathematics, numerical modeling, statistics, software engineering, and the ability to integrate data across multiple domains.
- Needs expertise in tools and techniques: rule learning, classification, cluster analysis, data fusion, machine learning, neural networks, anomaly detection, modeling, time series analysis, visualization.
- Needs 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.

Methodsology

1. Surveyed the curricula of 167 universities offering 267 ‘Data Analytics’ or ‘Data Science’ degree programs.
2. Categorized program based on focus area, by degree and on-line/off-campus.
3. Surveyed all curricula to determine course topics provided specific training, relevant to Earth science.
4. Charted course topics against program focus area to see what is emphasized per program.
5. Compare changes to 2013 survey.
6. Compare to ‘real world’ needs.

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.
- **Quantitative Analysis**: Database Management/Analysis, Artificial Intelligence.
- **Other Relevant Courses Offered**: Programming, Neural Networks, Data Analysis, Artificial Intelligence, Clustering, Time Series Analysis, Data Warehousing, Pattern Recognition, GIS, Remote Sensing, Text Mining, Information/Knowledge Management.

Summary of Best Practices

- Repeated Exposure
- Sharing of Vocabulary
- How, Where, When and Who to Find/Ask for Resources & Help
- Sharing of Community: “both ways” Communication.

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

<table>
<thead>
<tr>
<th>Program Focus Area</th>
<th>B on line</th>
<th>B campus</th>
<th>M on line</th>
<th>M campus</th>
<th>PhD online</th>
<th>PhD campus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business analytics</td>
<td>3</td>
<td>22</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data/Information science</td>
<td>15</td>
<td>10</td>
<td>21</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data analytics/Analytics</td>
<td>1</td>
<td>4</td>
<td>12</td>
<td>34</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Quantitative Analysis/Applied Statistics</td>
<td>5</td>
<td></td>
<td>17</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information technology/systems/management</td>
<td>4</td>
<td></td>
<td>19</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer science</td>
<td>1</td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health/Bio analytics</td>
<td>15</td>
<td></td>
<td>15</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business management</td>
<td>3</td>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental Informatics</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What Else Universities Should Consider Offering

- 1. Decreased for Business programs.
- 2. Increased in general Data Analytics/Science programs.
- 3. Decreased in Information technology/systems and Computer Science programs.
- 4. Added: a program called ‘Environmental 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.

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.
- **Quantitative Analysis**: Database Management/Analysis.
- **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 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**
- **Ability to integrate data across multiple domains**
- **Support domain scientists with data and computational needs to communicate across domains (be interdisciplinary)**
- **Knowledge of data lifecycle**
- **Software engineering - Programming**

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