



Tools for Building a **Big Data** Career Pathway

Joyce Malyn-Smith, *Principal Investigator*

Joe Ippolito, *Project Director*

Education Development Center



NSF-ATE DUE 1501927

Career Pathway Building Tools

The documents included here can assist schools looking to design and implement career pathways in big data. These tools have been used by community colleges involved in the NSF-ATE funded Creating Pathways for Big Data Careers (DUE-1501927). They consist of:

1. **Building Blocks for a Big Data Career Pathway**

This template can be used to present the programs, courses, supporting activities, school and employer linkages that comprise a school's proposed pathway.

2. **Gap Analysis Tool/ Industry Importance**

This tool enables local employers to indicate what big data skills they expect their workers to have and the level of proficiency they expect their workers to demonstrate for each of those skills. As presented here, the big data skills listed for analysis are those identified by a panel of big data experts who developed the occupational profile of a Data Practitioner.

3. **Gap Analysis Tool/ School Capability**

Using this tool, a school's faculty can identify the extent to which its current course offerings address the same big data skills and the depth to which those skills are addressed. Like the employer tool, this document presents for analysis the big data skills identified in the Data Practitioner profile. By comparing the employer and school analysis tools, faculty can identify "gaps" in their existing program offerings that may require modifications or additions to be made to their curricula. A comparison of the two tools may also reveal skills that currently receive excessive attention in a curriculum when compared to the importance employers place upon them.

4. **Employer Support Template**

Schools developing a career pathway can use this form in meetings with local employers to determine what kind of support they are willing to provide to a career pathway.

Further information on how to make use of these tools can be obtained by contacting:

Joyce Malyn-Smith, *Principal Investigator*, jmalynsmith@edc.org

Joe Ippolito, *Project Director*, jippolito@edc.org

NSF-ATE Creating Big Data Career Pathways Building Blocks for a Big Data Career Pathway

Please provide the following information regarding the programs, certificates and courses that relate to your school's development of a big data career pathway.

Your School _____

Person Completing Form _____

Two Year Programs

Name of Program	Indicate if – In place (X) or Planned (P), (with projected start date.)	Names of courses and course numbers	Prerequisites Necessary for Enrollment	Name(s) of Jobs Program Prepares Graduates For	Names of Companies Actively Supportive of Program

One Year Certificates

Name of Certificate	Indicate if – In place (X) or Planned (P), (with projected start date.)	Names of Courses Included and course numbers	Prerequisites Necessary for Enrollment	Name(s) of Jobs Certificate Prepares Graduates For	Names of Companies Actively Supportive of Certificate

Courses in Your School's Pathway

Name of Course	Indicate if – In place (X) or Planned (P), (with Projected start date.)	Name of Program/ Certificate	Brief Description	Number of Credits

Other courses and/ or activities not included above that relate to your school's big data career pathway.

Name of relevant Courses or Activity (e.g. internship)	How does this relate to other components of your career pathway?	Names of Companies Actively Supportive of Course or Activity

Identify Four Year Colleges that Your School has an Articulation Agreement with related to the program(s) and certificate(s) identified above.

Name of Program/ Certificate	Name of Four Year college that has articulation agreement with your school	Brief description of agreement	Is agreement already in place or in the planning stage?

Contact Information for Employers Supportive of Your School's Big Data Career Pathway

Name of Employer	Contact Person	Contact Information

GAP ANALYSIS TOOL/INDUSTRY IMPORTANCE
Creating Pathways for Big Data Careers

Name:

Organization:

Your Position:

Your e-mail:

When using this form, consider those individuals in your company who:

- Are data or data enabled workers
- Have 1-2 years experience
- Are Members of a data team
- Do not supervisor other data workers

INDUSTRY EXPECTATIONS	CODES	KEYWORDS
Not important for Data workers at my company	Not Present 0	
Data workers need to be aware of the knowledge/skill but are not required to use it.	Knowledge Level 1	Vocabulary Facts
Data workers are expected to understand, use and practice this knowledge/skill under supervision.	Practice Level 2	Understanding Use Do and practice Acts with help
Data workers are expected to demonstrate mastery of this skill, solve problems, integrate this with other skill/knowledge sets, apply this learning to new situations and teach others.	Mastery Level 3	Mastery Problem solving Integrate Teach others Complex Act independently Apply to new situations

(CTPAS, VISSP, and Education Development Center, Inc., 1996; adapted 2012)

Industry Partners:

*What expectations do you have for the performance of current **Data** workers on the following job functions?*

0=Not Present	1=Knowledge Level	2=Practice Under Supervision	3=Mastery Level
---------------	-------------------	------------------------------	-----------------

JOB FUNCTION	Performance Level
1. Initiates the Project	
2. Sources the Data	
3. Transforms the Data	
4. Analyzes the Data	
5. Closes out the Project	
6. Engages in Professional Development	

Industry Partners:

*What expectations do you have for the performance of current **Data** workers on the following tasks?*

0=Not Present	1=Knowledge Level	2=Practice Under Supervision	3=Mastery Level
----------------------	--------------------------	-------------------------------------	------------------------

DUTY 1: Initiates the Project

TASK	Performance Level
1A. Translates business problems into analytic needs.	
1B. Interviews stakeholders.	
1C. Refines stakeholder needs.	
1D. Identifies appropriate data.	
1E. Identifies whether data exists or not.	
1F. Performs gap analysis of the data.	
1G. Determines resource needs (e.g., SMEs, tools, timelines).	
1H. Determines feasibility of analysis to be done.	
1I. Creates statement of work.	

Industry Partners:

What expectations do you have for the performance of current **Data** workers on the following tasks?

0=Not Present	1=Knowledge Level	2=Practice Under Supervision	3=Mastery Level
---------------	-------------------	------------------------------	-----------------

DUTY 2: Sources the Data

TASK	Performance Level
2A. Determines data source(s)	
2B. Determines target structure.	
2C. Collects Data.	
2D. Exercises quality control (e.g., Randomizes selection).	
2E. Extracts data (e.g., writes SQL, API code).	
2F. Cleans data (e.g., identifies outliers/errors).	
2G. Tests data.	
2H. Creates data dictionary.	
2I. Complies with business, ethical and legal standards.	

Industry Partners:

What expectations do you have for the performance of current **Data** workers on the following tasks?

0=Not Present	1=Knowledge Level	2=Practice Under Supervision	3=Mastery Level
---------------	-------------------	------------------------------	-----------------

DUTY 3: Transforms the Data

TASK	Performance Level
3A. Merges data.	
3B. Splits data.	
3C. Derives new variables.	
3D. Creates new data.	
3E. Augments data.	
3F. Applies metadata.	
3G. Purges data.	
3H. Changes data structure.	
3I. Changes data types.	

3J. Normalizes data.	
3K. Interpolates data.	
3L. Finalizes data dictionary.	
3M. Stores data for analytics.	

Industry Partners:

What expectations do you have for the performance of current **Data** workers on the following tasks?

0=Not Present	1=Knowledge Level	2=Practice Under Supervision	3=Mastery Level
---------------	-------------------	------------------------------	-----------------

DUTY 4: Analyzes the Data

TASK	Performance Level
4A. Determines what analysis to run.	
4B. Applies the research method and tools.	
4C. Identifies dependent and independent variables.	
4D. Defines appropriate algorithms.	
4E. Performs data mining.	
4F. Separates any anomalies.	
4G. Interprets the results.	
4H. Runs additional tests as needed.	
4I. Performs reasonableness tests of results.	

4J. Compares results to previous findings.	
4K. Confirms results.	
4L. Conducts causality testing.	
4M. Creates data visualizations (e.g., dashboards, reports, charts, graphs, videos, animation).	

Industry Partners:

What expectations do you have for the performance of current **Data** workers on the following tasks?

0=Not Present	1=Knowledge Level	2=Practice Under Supervision	3=Mastery Level
---------------	-------------------	------------------------------	-----------------

DUTY 5: Closes Out the Project

TASK	Performance Level
5A. Selects documentation media.	
5B. Describes problem, method and analysis.	
5C. Articulates conclusions.	
5D. Complies reports.	
5E. Presents information to stakeholders.	
5F. Integrates feedback from stakeholders.	
5G. Defends analysis as needed.	
5H. Reworks analysis as needed.	
5I. Prepares final report.	

5J. Archives work products.	
5K. Communicates future processes, improvements and opportunities.	

Industry Partners:

*What expectations do you have for the performance of current **Data** workers on the following tasks?*

0=Not Present	1=Knowledge Level	2=Practice Under Supervision	3=Mastery Level
---------------	-------------------	------------------------------	-----------------

DUTY 6: Engages in Professional Development

TASK	Performance Level
6A. Maintains professional qualifications.	
6B. Stays current on emerging technologies, methods and tools.	
6C. Seeks out mentors.	
6D. Shares best practices.	
6E. Contributes new knowledge to the field.	
6F. Attends relevant conferences and seminars.	
6G. Mentors others.	
6H. Participates in professional organizations.	
6I. Suggests future projects.	

Having reviewed in detail the tasks involved in being a Data Practitioner, and considering the previous dialog, please answer the following question:

What job (s) at your company require the use of all or most of the tasks you have just analyzed?

1. Title of Job: _____

2. Title of Job: _____

3. Title of Job: _____

GAP ANALYSIS TOOL \ SCHOOL CAPABILITY
Creating Pathways for Big Data Careers

Name: _____

Course/ Program: _____

Email: _____

When using this form, consider the current status of curriculum as related to each performance objective

INSTITUTIONAL CAPABILITY	CODES	KEYWORDS
Not covered in curriculum	Not Present 0	
Curriculum introduced through vocabulary, readings and facts.	Knowledge Level 1	Vocabulary Facts
Curriculum provides opportunities to understand, use and practice this knowledge/skill, with the help of an instructor.	Practice Level 2	Understanding Use Do and practice Acts with help
Curriculum provides opportunities to master the skill, solve problems, integrate, act independently, apply this learning to new situations and teach others.	Mastery Level 3	Mastery Problem solving Integrate Teach others Complex Act independently Apply to new situations

(CTPAS, VISSP, and Education Development Center, Inc., 1996; adapted 2012)

Faculty:

To what extent does our current curriculum address the performance of these objectives?

0=Not Present	1=Knowledge Level	2=Practice, With Help	3=Mastery Level
----------------------	--------------------------	------------------------------	------------------------

DUTY	Performance Level
1. Initiates the Project	
2. Sources the Data	
3. Transforms the Data	
4. Analyzes the Data	
5. Closes out the Project	
6. Engages in Professional Development	

Faculty:

To what extent does our current curriculum address the performance of these objectives?

0=Not Present

1=Knowledge Level

2=Practice, With Help

3=Mastery Level

DUTY 1: Initiates the Project

TASK	Performance Level
1A. Translates business problems into analytic needs.	
1B. Interviews stakeholders.	
1C. Refines stakeholder needs.	
1D. Identifies appropriate data.	
1E. Identifies whether data exists or not.	
1F. Performs gap analysis of the data.	
1G. Determines resource needs (e.g., SMEs, tools, timelines).	
1H. Determines feasibility of analysis to be done.	
1I. Creates statement of work.	

Faculty:

To what extent does our current curriculum address the performance of these objectives?

0=Not Present	1=Knowledge Level	2=Practice, With Help	3=Mastery Level
---------------	-------------------	-----------------------	-----------------

DUTY 2: Sources the Data

TASK	Performance Level
2A. Determines data source(s)	
2B. Determines target structure.	
2C. Collects Data.	
2D. Exercises quality control (e.g., Randomizes selection).	
2E. Extracts data (e.g., writes SQL, API code).	
2F. Cleans data (e.g., identifies outliers/errors).	
2G. Tests data.	
2H. Creates data dictionary.	
2I. Complies with business, ethical and legal standards.	

Faculty:

To what extent does our current curriculum address the performance of these objectives?

0=Not Present

1=Knowledge Level

2=Practice, With Help

3=Mastery Level

DUTY 3: Transforms the Data

TASK	Performance Level
3A. Merges data.	
3B. Splits data.	
3C. Derives new variables.	
3D. Creates new data.	
3E. Augments data.	
3F. Applies metadata.	
3G. Purges data.	
3H. Changes data structure.	
3I. Changes data types.	

3J. Normalizes data.	
3K. Interpolates data.	
3L. Finalizes data dictionary.	
3M. Stores data for analytics.	

Faculty:

To what extent does our current curriculum address the performance of these objectives?

0=Not Present

1=Knowledge Level

2=Practice, With Help

3=Mastery Level

DUTY 4: Analyzes the Data

TASK	Performance Level
4A. Determines what analysis to run.	
4B. Applies the research method and tools.	
4C. Identifies dependent and independent variables.	
4D. Defines appropriate algorithms.	
4E. Performs data mining.	
4F. Separates any anomalies.	
4G. Interprets the results.	
4H. Runs additional tests as needed.	
4I. Performs reasonableness tests of results.	

4J. Compares results to previous findings.	
4K. Confirms results.	
4L. Conducts causality testing.	
4M. Creates data visualizations (e.g., dashboards, reports, charts, graphs, videos, animation).	

Faculty:

To what extent does our current curriculum address the performance of these objectives?

0=Not Present

1=Knowledge Level

2=Practice, With Help

3=Mastery Level

DUTY 5: Closes Out the Project

TASK	Performance Level
5A. Selects documentation media.	
5B. Describes problem, method and analysis.	
5C. Articulates conclusions.	
5D. Complies reports.	
5E. Presents information to stakeholders.	
5F. Integrates feedback from stakeholders.	
5G. Defends analysis as needed.	
5H. Reworks analysis as needed.	
5I. Prepares final report.	

5J. Archives work products.	
5K. Communicates future processes, improvements and opportunities.	

Faculty:

To what extent does our current curriculum address the performance of these objectives?

0=Not Present

1=Knowledge Level

2=Practice, With Help

3=Mastery Level

DUTY 6: Engages in Professional Development

TASK	Performance Level
6A. Maintains professional qualifications.	
6B. Stays current on emerging technologies, methods and tools.	
6C. Seeks out mentors.	
6D. Shares best practices.	
6E. Contributes new knowledge to the field.	
6F. Attends relevant conferences and seminars.	
6G. Mentors others.	
6H. Participates in professional organizations.	
6I. Suggests future projects.	

Employers Supporting (Name of School's) Big Data Career Pathway

Please indicate all actions you and/ or your company are willing to take to support (Name of School's) Big Data Career Pathway.

Name _____

Position _____

Company _____

Telephone _____

E-Mail _____

Supportive Action	Yes!	Maybe, more discussion needed	No
Provide input into curriculum development			
Provide guest speakers for data courses			
Host field trips by students enrolled in data courses			
Allow students to "shadow" data workers			
Offer unpaid internships			
Offer paid internships			
Conduct mock job interviews			
Participate in college sponsored career days			
Serve on Industry Advisory Committee			
Provide labor market data, e.g. info about job openings			
Support fundraising efforts to strengthen the pathway			
Other?			
Other?			