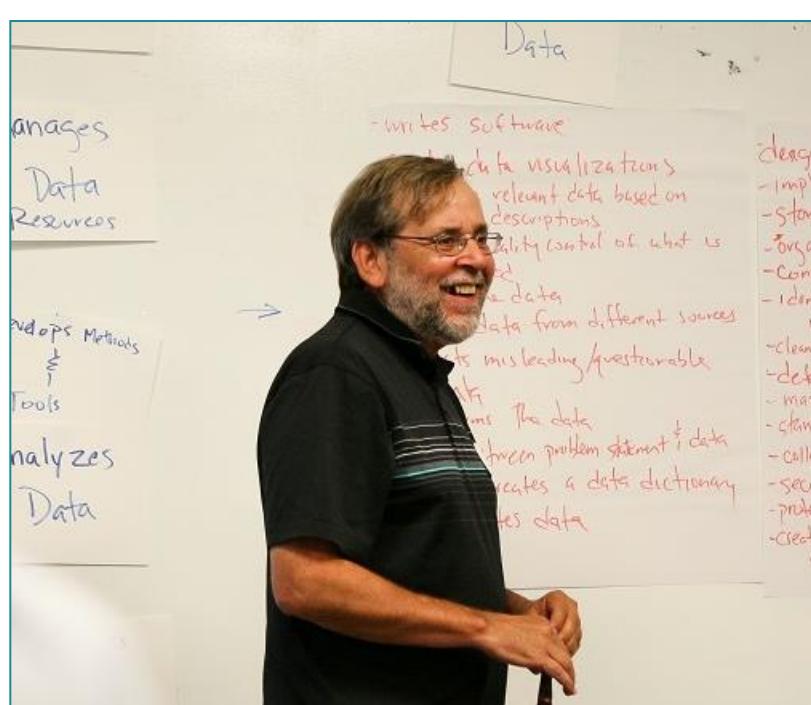


DEVELOPING AN OCCUPATIONAL SKILLS PROFILE FOR THE EMERGING PROFESSION OF “BIG-DATA-ENABLED PROFESSIONAL”

Kim Kastens, Joyce Malyn-Smith, Joseph Ippolito, Ruth Krumhansl (for more information, please contact Ruth Krumhansl at oceansofdata@edc.org)
Poster ED31E-3458 – Presented at American Geophysical Union, December 17, 2014

THE PROCESS

EDC's approach integrates two methods used to develop curriculum for technical training programs.



EDC's modified DACUM methodology rests upon three basic principles:

- Expert workers can describe and define their jobs more accurately than anyone else.
- An effective way to define a job is to precisely describe the tasks that expert workers perform.
- All tasks, in order to be performed correctly, demand certain knowledge, skills, resources, and behaviors.

DACUM Steps:

1. A panel of 8-12 expert workers is recruited, representing a range of occupational levels and work settings.
2. EDC works with the panel on-line to draft an initial occupational definition.
3. The panel is convened for a two-day intensive work session.
4. The profile is validated by a larger community of data scientists using a customized on-line occupational survey.



USES

1. In education, for designing curricula, programs, assessments, and certifications.
2. In the workforce, for developing job descriptions, mentoring data scientists, and designing professional development programs.

THE PRODUCT

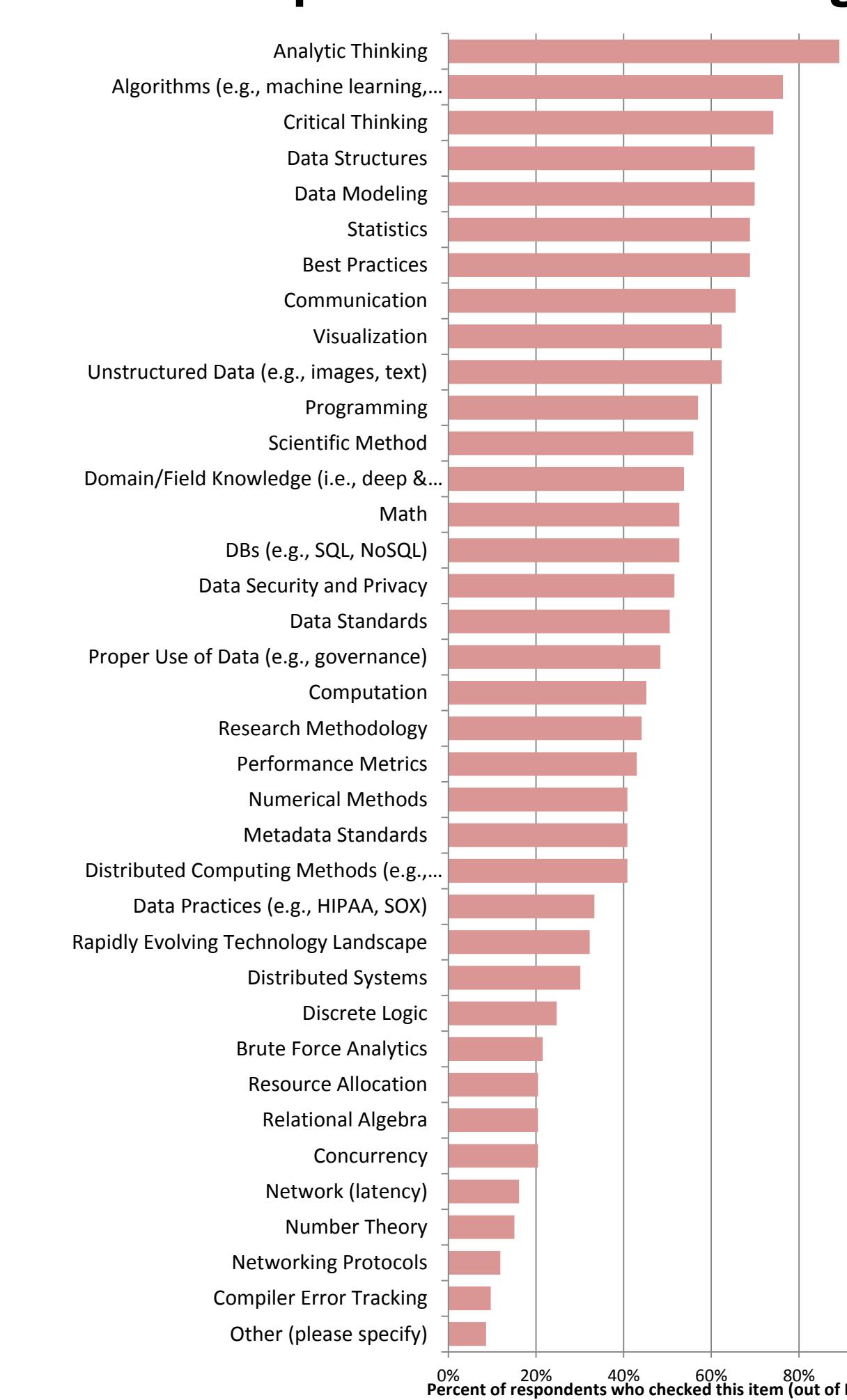
Definition: The big-data-enabled specialist is an individual who wrangles and analyzes large and/or complex data sets to enable new capabilities including discovery, decision support, and improved outcomes.

DUTIES	TASKS										
	1A. Identifies stakeholders	1B. Determines stakeholders' needs	1C. Articulates the question	1D. Aligns study to organizational goals and objectives	1E. Translates question into a research plan	1F. Designs the experiment	1G. Develops deep domain knowledge of data source	1H. Discerns whether big data is needed to solve problem	1I. Identifies resources (e.g., experts, software)	1J. Performs gap analysis	1K. Assesses risk and bias involved in conducting study/project
1. Defines the Problem	1L. Communicates costs/risks of study to stakeholders	1M. Negotiates plan, including deadlines and budgets	1N. Creates requirement document (sign-off)								
2. Wrangles Data	2A. Performs data exploration	2B. Identifies data	2C. Creates the data dictionary	2D. Collects data	2E. Assesses the extent/methods to clean the data	2F. Maps data across heterogeneous sources	2G. Identifies outliers and anomalies	2H. Cleans data	2I. Transforms data	2J. Synthesizes data	2K. Defines new metrics/attributes based on accessible data
3. Manages Data Resources	2L. Performs data visualization	2M. Writes software to automate tasks	2N. Documents the process								
4. Develops Methods and Tools	3A. Manages data life cycle	3B. Conducts capacity planning of resources	3C. Complies with legal obligations	3D. Applies ethical standards	3E. Identifies tools that may be needed for purchase or modification	3F. Protects data and results	3G. Determines access to the data	3H. Designs ETL workflow	3I. Implements ETL workflow	3J. Stores the data	3K. Upserts data sources
5. Analyzes Data	4A. Researches current methods/models	4B. Extends existing methods/models, if possible	4C. Selects tools/software/programming environment	4D. Develops new methods/models	4E. Runs simulations	4F. Iterates correctness and scalability of methods/models	4G. Validates methods/models with test cases	4H. Disseminates methods/models for peer review	4I. Documents methods/models		
6. Communicates Findings	5A. Develops analysis plan	5B. Applies methods and tools	5C. Conducts exploratory analysis (e.g., identifies anomalies, outliers, bias in sampling; visualizes)	5D. Evaluates results of the analysis (e.g., significance, effect, size)	5E. Estimates precision and accuracy of answer	5F. Determines level of confidence in results	5G. Compares results with other findings	5H. Answers the question (e.g., insights drawn from results)	5I. Submits preliminary findings for peer review	5J. Documents preliminary findings	
7. Engages in Professional Development	6A. Selects documentation media (e.g., dashboard, PowerPoint, e-mail)	6B. Compiles report	6C. Describes the problem, method, and analysis	6D. Identifies limitations (e.g., data use, data application methods)	6E. Scopes data narrative based on time, depth, and method	6F. Prepares visualizations	6G. Guides interpretation	6H. Articulates conclusions	6I. Contrasts alternative approaches and past results	6J. Provides recommendations based on results	6K. Tells “data story” to convey insight (e.g., talks to CEO)
	7A. Seeks out mentors	7B. Stays current on emerging technologies, data types, and methods	7C. Attends relevant big data conferences	7D. Contributes new knowledge to the field	7E. Maintains professional library	7F. Participates in professional organizations	7G. Mentors others	7H. Engages in cross-discipline training	7I. Articulates value of big data activities to other departments/functions of the organization	7J. Articulates evolving role of big data in supporting organizational goals	

VALIDATION

One hundred additional big-data-enabled professionals from 15 industry sectors completed an extensive online survey.

KNOWLEDGE: An effective big-data-enabled specialist has knowledge of:



SKILLS: An effective big-data-enabled specialist possesses skills in:



BEHAVIORS: A successful big-data-enabled specialist is:

