THE UNIVERSITY of TENNESSEE KNOXVILLE

BIG ORANGE BIG IDEAS

Knowledge, skills, and abilities of earth science data job incumbents

Wade Bishop, Ph.D.
School of Information Sciences
University of Tennessee

April 28, 2017
Belmont Forum Digital Skills and Curricula Development Workshop

Overview

- This study's purpose was to explore the job practices of earth science data managers as they relate to the sequential steps of the data lifecycle.
- Twelve earth science data managers were interviewed using a Developing a Curriculum (DACUM) approach, which focuses on job tasks and their frequencies as explained by current workers.
- Several actions related to the data lifecycle, such as data discovery, do require an understanding of the data, technology, and information infrastructures that may result from information science education.

Developing a Curriculum (DACUM)

- The (Developing a Curriculum) DACUM approach creates a list of knowledge, skills, and abilities, operationalized job descriptions, and eventually learning outcomes for use in education and training for jobs.
- The DACUM approach builds on two assumptions. The first assumption is that current workers know how to do their jobs and should be the best at describing them (Knapp & Knapp, 1995).
- In emerging fields, it is key to study Subject Matter Experts (SMEs) as they are creating the profession through their practice.

Challenges and Approaches

- Ideally, the researcher conducing a DACUM should reduce biases by both sampling workers evenly across a profession, as well as having little to no experience with the job.
- The approach includes focus groups or interviews with SMEs to discover trends in job practices, then validating a list of knowledge, skills, and abilities through a survey of more SMEs to lead to representative results.
- The outcome of these iterative research steps is a chart, which lists the knowledge, skills, and abilities needed to work in a profession.

Interview questions

1) What is your current job title?
2) How many years in total have you been working in your current job?
3) How many years in total have you been working with earth science data?
4) Describe your work setting?
5) Please indicate your credentials and degrees.
6) Please provide any other education or training you have received that is applicable to performing your job.
7) What are some daily tasks associated with the job?
8) What are some weekly tasks associated with the job?
9) What are some less frequent tasks associated with the job?
10) What are some tasks NOT associated with the job?

Diversity in Job Titles and Degrees

- Data Manager (2), Data Curation and Stewardship Coordinator, Branch Chief, Software engineer, Director of a statewide Geologic Survey, scientific programmer, Data Scientist Level 2, Technical Project Manager and Software Engineer, Biologist ("but my boss considers me a science data manager"), Information Architect and Principal Computer Scientist, Student, and Geographic Information Systems (GIS) Analyst
- ~16 years experience with earth science (includes higher education).
- Workplace longevity range of a few months to 20 years with an average of about ~7 years.

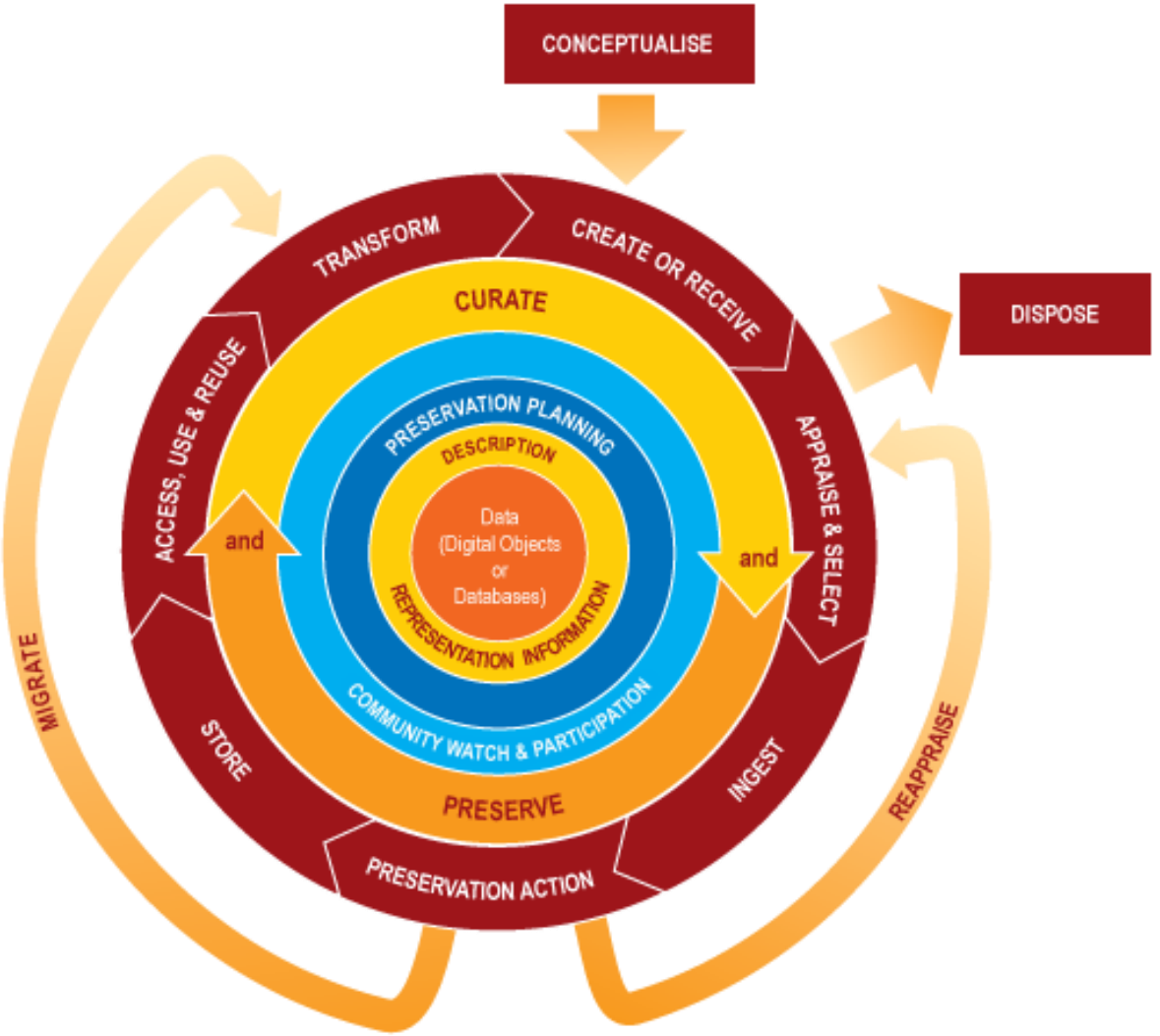
Diversity in Degrees

- All twelve participants held bachelor's degrees and those included, Computer Science (4), Biology (2), Physics (2), Electrical Engineering, Geology, Geography, Meteorology, and English.
- Ten of the twelve held master's degrees, including Library and Information Science (2), Computer Science, Wildlife Biology, Conservation Biology, Atmospheric Science, Geography (specialized in Atmospheric science), Weather Climate and Modeling, Electrical Engineering, Mechanical Engineering, Geology and Geophysics.
- One participant in defense of having only a bachelor's degree said "I came out of school sort of at the right time for someone in this field. You didn't have to have a Master's or a Ph.D. to get into doing really serious stuff."
- Three participants did have Ph.D.s, with one each in Geology and Geophysics, Ecology, and Atmospheric Sciences.

Training and software mentioned

- To perform these jobs, participants mentioned attending hands-on trainings, viewing online course modules, consulting informal sources, reading papers and workflows, and listed a few specific resources; DataONE-data curation or management, Globe.gov, Coursera courses, and Code academy.
- Participants consulted all these materials with the intentions of "catching up" on new programming languages, software tools and frameworks, data carpentry, metadata standards, and information modeling.
- The tools, formats, and other items mentioned by name from the trainings mentioned included, Python (5), NetCDF (5), R (3), C, Fortran, Java script (2), XML (2), Mircosoft Outlook 365 (2), Anaconda, Pearl, C+, CSS, ArcGIS, Cold Fusion, Hadoop, Spark, SciSpark, Scrum, Protégé, Oxygen XML, Altova MissionKit, Eclipse Integrated Development Environment (IDE), PuTTY, Quantitative Insights into Microbial Ecology (QIME), Hadoop, EmEx, Metadata Parser, and other metadata creation tools that mint DOIs.

UK Digital Curation Centre (DCC) Curation Lifecycle Model



Job-related tasks through the Lifecycle

1. Create or receive data, including administrative, descriptive, structural and technical metadata creation
2. Appraisal and select
3. Ingest (final aspects of data quality assurance at the point of deposit)
4. Preservation Action (data remains authentic, reliable and usable while maintaining its integrity)
5. Store
6. *Access, Use, and Reuse*
7. Transform (e.g., migration into a different format or creating a subset)

Preservation Tasks

- Create or receive data
  - Metadata compliance (2); incomplete/incompatible metadata (2); CSDGM (3); Dublin Core; ISO; “metadata discovery”; talk to people about their data
- Appraisal and select
  - submitting initial metadata about their data set, so that the research data archive can make an informed decision whether or not the data submission is aligned with our mission statement
  - Remind scientists to turn in their data “PI wrangling”
- Ingest (final aspects of data quality assurance at the point of deposit)
  - Quality assurance automation (2)
  - Testing data submission process
- Preservation Action (data remains authentic, reliable and usable while maintaining its integrity)
  - Evaluate best practices and how they may benefit a repository
- Store
  - DAC or node; long-term repository; physical samples

Curation Tasks

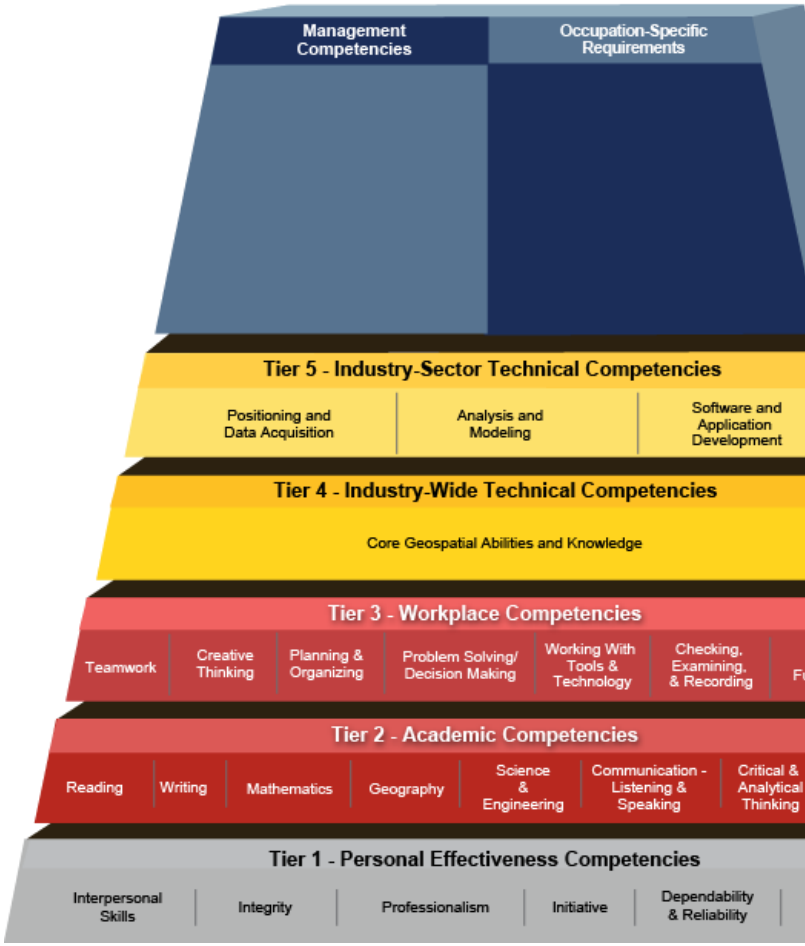
- *Access, Use, and Reuse*
  - Question Answering (8)
  - Usability/design (2)
  - "So we built this tool, our repository, and it's not always very intuitive to people, so we're on the phone, 'okay click this button' or 'how do I write my metadata'."
- Transform (e.g., migration into a different format or creating a subset)
  - Compiling data to move to another format
  - Create maps and 3D images

Some quotes

- "you've got to actually be able to program, you've got to be able to do that back end stuff"
- "I want to see domain scientists also exposed to some basic data management practices so that they understand that there might be somebody around that can help them with that instead of reinventing their own wheel"
- "You need to understand data. You need to understand science. You need to understand technology. You need to understand infrastructures. Those four things you need to at least pull together."
- "it would be good for scientists of all stripes to have more experience with data, how data works, where it comes from, how it is stored, what are the problems with data"
- "It seems crazy to me of the required coursework for an Ecology student that, of the things that are there, it is all theoretical. There is not a single piece of computing or data science that is incorporated as an explicit requirement in the degree. Considering that is probably the most important piece in the whole degree in many ways, it seems silly to me."

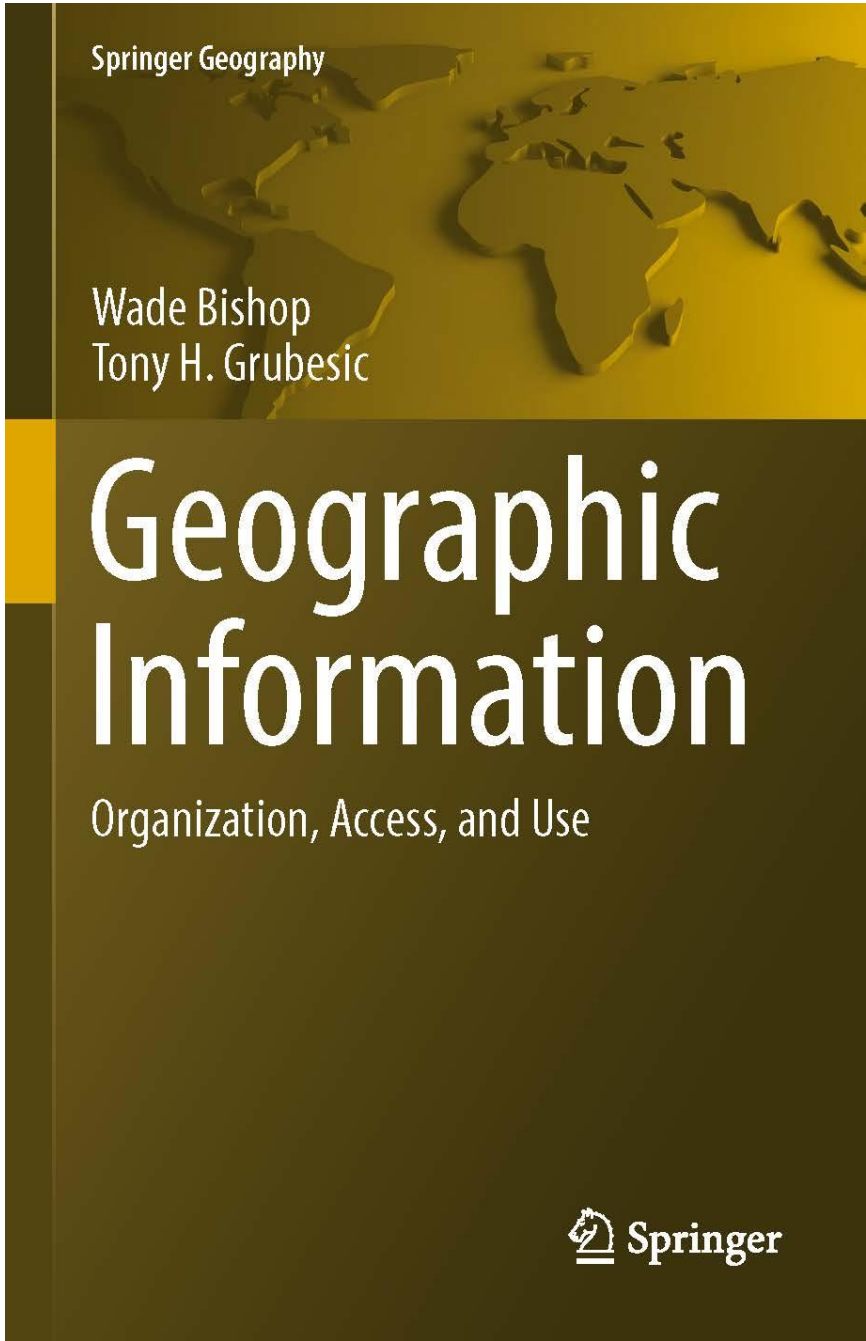
Next steps

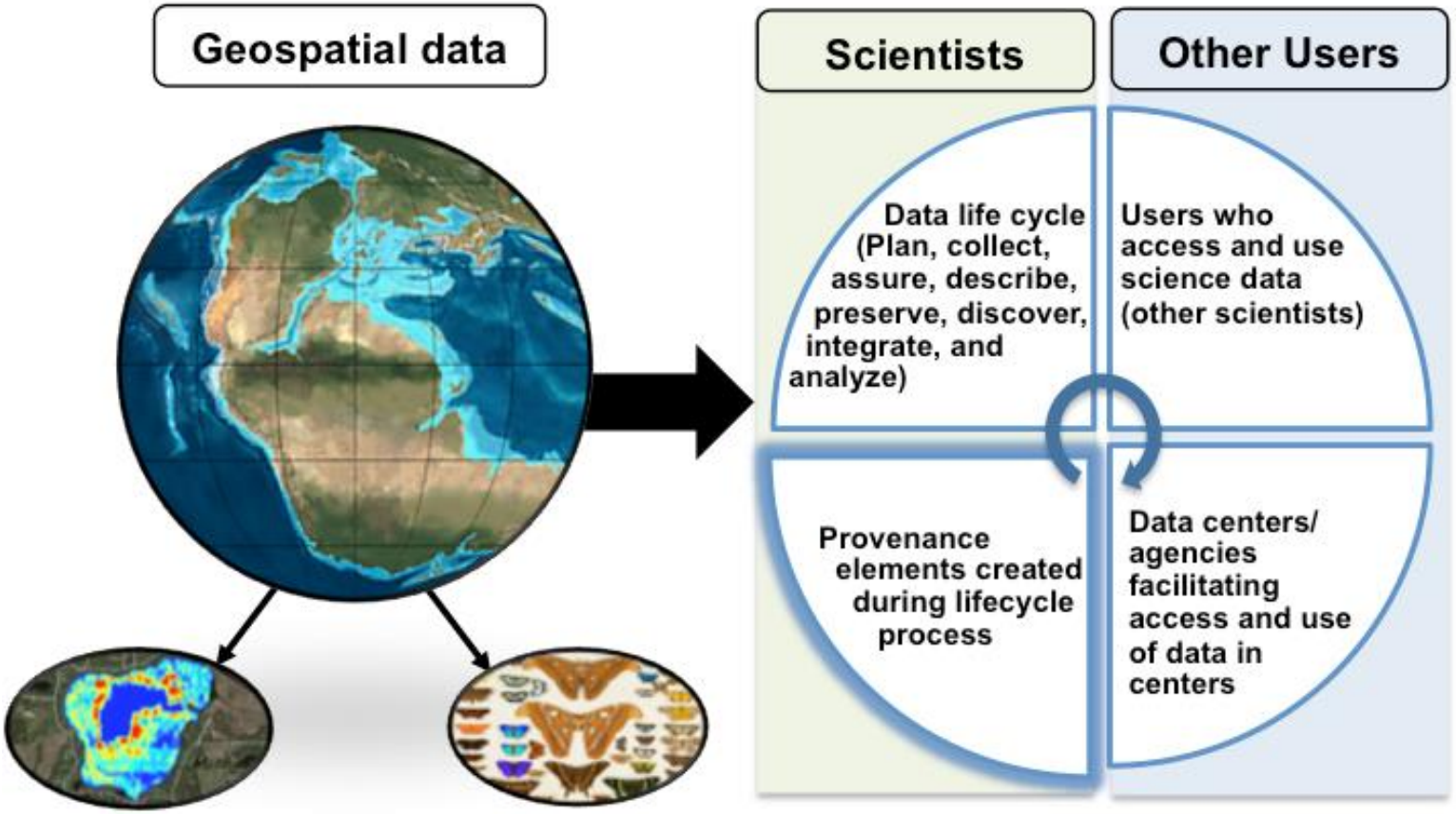
- A survey of more earth science data managers across sectors and regions could produce representative results and likely allow for the discovery of other tasks performed.
- A DACUM chart that lists the knowledge, skills, and abilities needed to work in this profession would help steer continuing education frameworks for both data producers and curators.
- These results may also help address the need for more core science data curricula to address the workforce development issue of a shortage of data curators.



GI: Organization, Access, and Use

- Bishop, B.W. & Grubesic, T. H. (2016). *Geographic Information: Organization, Access, and Use of Geographic Information*. Springer.





Provenance data research to increase usability and accessibility, data science undergraduate education, and 9-12 teacher outreach

Comments, Questions

wade.bishop@utk.edu