



Data Skills & Competencies Requirements for Data Specialists:

Views from the ESIP Community & Beyond

Nancy Hoebelheinrich (Knowledge Motifs LLC) &
Karl Benedict (University of New Mexico)



This project is made possible by a grant from the U.S. Institute of Museum and Library Services (IMLS). IMLS Grant LG-70-18-0092-18. The DMTC is managed by the Earth Science Information Partners, with development support provided by DataONE, and USGS. The views, findings, conclusions or recommendations expressed in this presentation do not necessarily represent those of the IMLS, or the other sponsor organizations.

<http://bit.ly/>

Presentation Outline

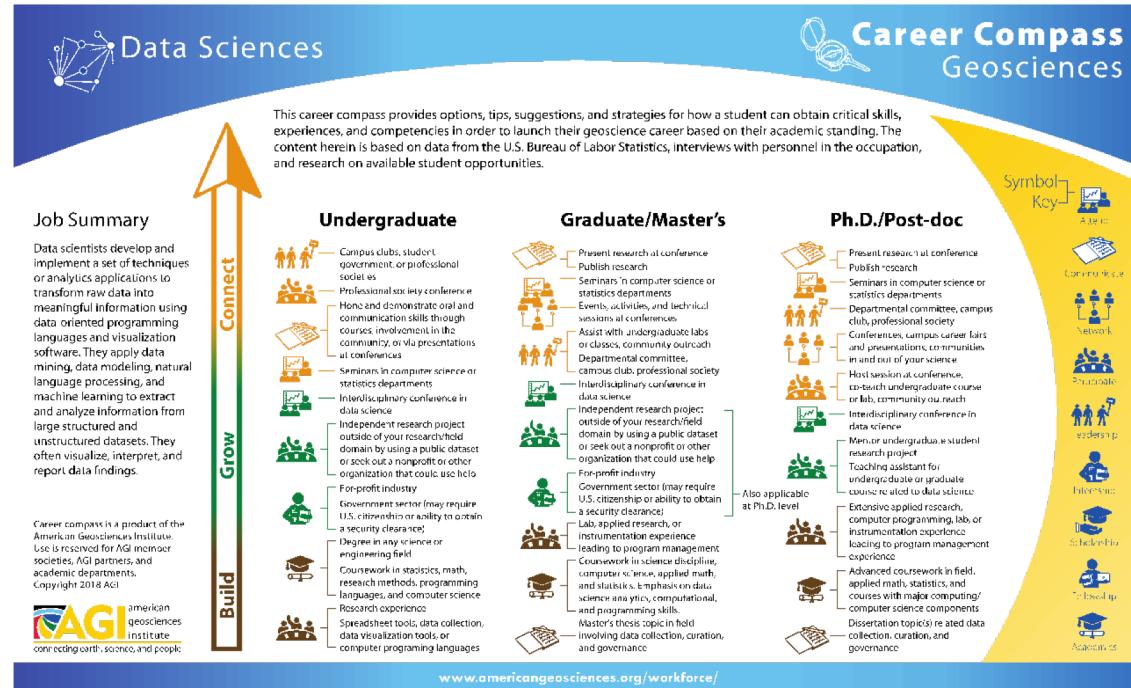


1. Session Goals & Objectives
2. Introduction to a skills & capabilities approach
3. Introduction to assessing capabilities with data lifecycle stages in mind, from the POV of the research object
4. Your feedback requested via a survey monkey form.

Image from Shutterstock

Goals & Objectives

- Identify core knowledge, skills, and competencies required / desirable for data “specialists” from an Earth Science POV
- Contribute an American Geosciences Institute Career Compass for Data Stewards
- Inform the assessment work of the Data Management Training Clearinghouse



further test what the EOSC project came up with against Earth Science “data advisors” and “data service providers”

Adapted from Whyte, A., et al, "Annex B: Skills Tables" of "D7.5 Strategy for Sustainable Development of Skills and Capabilities", EOSC pilot, v1.1, April 29, 2019.
<https://eoscpilot.eu/content/d75-strategy-sustainable-development-skills-and-capabilities>. Last viewed September 17, 2019.

Step 1: Interactive Poster from ESIP Summer 2019



Interactive Poster results in spreadsheet form

Task: To identify knowledge, skills, and competencies for data “advisors” and data “service providers”

Areas of Responsibility	Competencies	Data "Advisors" (e.g., "curators")				Data "Service Providers" (e.g., "data librarians")			
		High	Medium	Low	Not applicable	High	Medium	Low	Not applicable
Awareness & Support for Managing data	Context (i.e. 1) research lifecycles, 2) data lifecycles, 3) data management standards, & 4) roles involved in research & data management). [Row 1]	9	2		1			6	
	Activities to support data access and reuse . [Row 2]	4	6			4	2	2	
	Data and metadata best practices (i.e. 1) data structures, types, formats, vocabularies), 2) metadata, and 3) ontologies). [Row 3]	10	1	1		9	1		
	Function of data management plans and tools [Row 4]	10	1	1			3	3	1
	Scholarly publication requirements & role of open access. [Row 5]	5	6				1	3	2
	Data sharing options including IPR & licensing options. [Row 6]	3	3	2			2	2	1
	Discipline specific funders' policies & requirements . [Row 7]	4	3	3		1	3	2	1
	Domain / discipline / institution specific data repositories' requirements . [Row 8]	8	2	1		2	5		
Managing Data Collections	Data collection evaluation & assessment for retention as well as immediate and ongoing data management needs. [Row 9]	1	5	1		5		4	
	Data linking/integration/discovery techniques & tools . [Row 10]	4	5			8	1		
	Applications/adaptions of standards , including metadata schemas, data formats, domain ontologies, identifiers, data citation, and data licenses. [Row 11]	4	1	3		6			
	Data repository and storage platforms , database design types and structures. [Row 12]	1	5	3		9	1		
Teaching & Training	Data wrangling techniques and tools for tasks related to data visualization, programming (such as R, Python, Javascript, etc.), website development and maintenance, application of statistical software (SAS, MATLAB, SPSS, etc.), and GIS software. [Row 13]		6	3	1	9			
	Consultation and training development , including online tutorials, course materials or instructional guides. [Row 14]	9	1			3	2	3	
	Integration of educational & training into full courses, and with course, instructional program and service evaluation. [Row 15]	1	6	3		2	3	2	1
Outreach & Networking	Collaborative relationship development with other research and data service team members. [Row 16]	7	3			3	4		
	Effective communication through a variety of techniques including scholarly & more informal publication, oral presentations and small group discussion, and social media. [Row 17]	11	2			3	4	1	

What we learned: Important for responders to better understand the difference b/w roles associated with Data “Advisors” vs. “Service Providers”

Recommended expertise by professional group and service role									
Data Advisors					Data Service providers				
Data Steward	Research Mgr	Ethics & DP	Comm. & IP	Data Librarian	Research Swr Eng	Service Mgr	Service Architect	Archivist	
●	●	●	●	●	●	○	○	○	

EOSC report, Table 12, pg. 80

Brief Role Profiles: Key Tasks (from enabling FAIR data perspective)

From EOSC Report, Annex B, B.4.3

Data Service Provider: Data Librarian (pgs .73-74)

- **Use or develop FAIR research tools or services**
- **Apply policies to comply with legal, ethical and FAIR principles**
- **Prepare and document data/code to make outputs FAIR**
- **Develop open research strategy and vision**

NOTE: Institutional Level Librarian

Data Advisor: Data Steward (pgs. 75-6)

- **Plan for the stewardship and sharing of FAIR outputs**
 - **Use or develop FAIR research tools or services**
 - **Prepare and document data/code to make outputs FAIR**
 - **Publish FAIR outputs on recommended repositories**
- NOTE: **Operational Level position**, e.g., research group or department / faculty

further test what the EOSC project came up with against Earth Science “data advisors” and “data service providers”

Adapted from Whyte, A., et al, "Annex B: Skills Tables" of "D7.5 Strategy for Sustainable Development of Skills and Capabilities", EOSC pilot, v1.1, April 29, 2019.
<https://eoscpilot.eu/conten/d75-strategy-sustainable-development-skills-and-capabilities>. Last viewed September 17, 2019.

Step 2: Small group assessment of competencies according to data lifecycle stages using **DataONE's data lifecycle**



Work in process

Point of View:

We are assessing competencies and how they apply to a data lifecycle stage from the **POV of the research object itself**

Premises:

- Allows us to look from ***the data perspective*** rather than from that of either a data advisor or a data service provider when deciding to which part(s) of the lifecycle a given competency should relate.
- To scope, we assess which are the most important /relevant steps of the lifecycle by judging which have ***the highest impact*** on the research object given the category of activity (i.e., plan and design; capture and process, etc.) in which the competency falls.

Survey: <http://bit.ly/data-competencies>

Work in process: Example

FAIR4S competences and capabilities		Steps of DataONE Data Lifecycle									Comments
		Plan	Collect	Assure	Describe	Preserve	Discover	Integrate	Analyze		
<i>From perspective of the RO; evaluating the tasks to determine at which stage in the data lifecycle they are most applicable.</i>											
Capture and process	Appraise and select repositories for FAIR sharing	x			?	x	x				
	Reuse data from existing sources										
	Manage databases		x	x					x		
	Software prototyping			x	x				x		Includes documentation of the plan & processes desired for the SW.
	Set up and document workflows	x		x	x	x					Added plan as there was no equivalent for planning of workflows in the Plan and Design categories.
	File naming and organisation	x		?	x					?	Assure added as the most impactful effect in terms of scale of the workflows as opposed to Integrate & Analyze.
	Data cleaning, processing and software versioning			x				x	x		Scheme needed to be able to distinguish b/w source data & product data.

https://docs.google.com/spreadsheets/d/1F9OX6z3ZnZw2kdigibMERW_ZhHi2jN49JPOXwoZrPo8/edit?usp=sharing

Work in process: Discussion & invitation to join the process

https://docs.google.com/spreadsheets/d/1F9OX6z3ZnZw2kdi9ibMERW_ZhHi2jN49JPOXwoZrPo8/edit?usp=sharing

further test what the EOSC project came up with against Earth Science “data advisors” and “data service providers”

Step 3: Community feedback via Survey Monkey form asking for evaluation similar to interactive poster, but rating competencies from a range of “fluency” to “novice”.

Survey: https://www.surveymonkey.com/r/2020-01_ESIP_Competencies

Adapted from Whyte, A., et al, "Annex B: Skills Tables" of "D7.5 Strategy for Sustainable Development of Skills and Capabilities", EOSC pilot, v1.1, April 29, 2019.
<https://eoscpilot.eu/conten/d75-strategy-sustainable-development-skills-and-capabilities. Last viewed September 17, 2019.>