

Sort of... but not really.

Research Overview

- Penn State University - R1 university
 - 24 unique campuses (23 physical, 1 virtual)
 - 5 campuses with graduate students
- STEM disciplines are among highest enrolled programs
- STEM research expenditures highest at campuses with highest enrollments
 - Not all campuses with STEM programs offer research opportunities/funding
- Conducted quantitative survey and followed up with qualitative focus groups

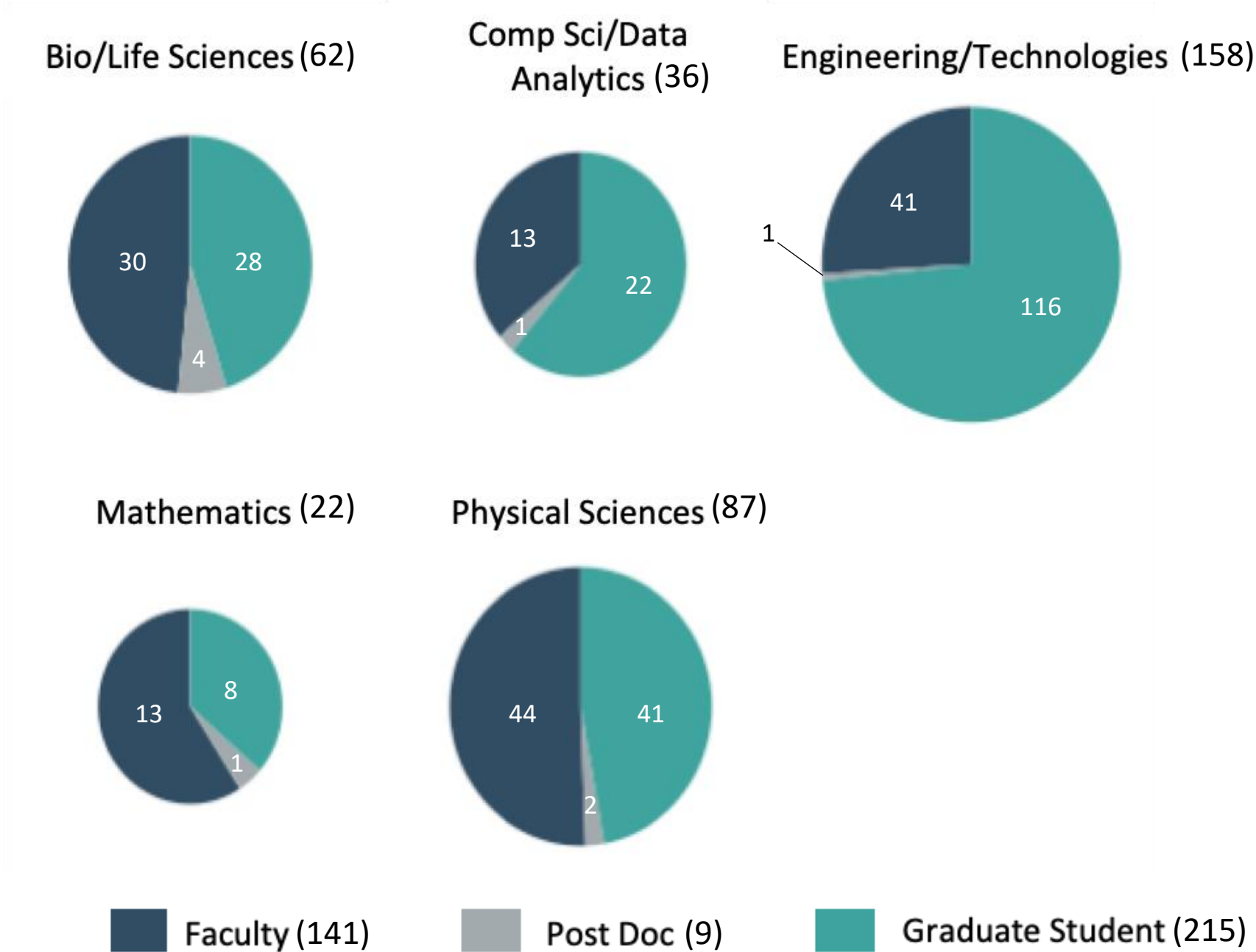
Research Purpose & Goals

- Gain a more in depth understanding of research data management awareness and practices of STEM researchers across various PSU campuses
- Improve research reproducibility and support open science initiatives by improving our understanding of the RDM ecosystem at PSU
- Assist us (University Libraries) in identifying new ways to support STEM research at PSU

Research Questions

1. Does awareness of various RDM practices and services differ by campus, career stage, or discipline?
2. Do current RDM practices differ by campus, career stage, or discipline?
3. Do researchers understand the terminology commonly employed by librarians when discussing RDM?
4. What terminology do researchers from different disciplines use when discussing RDM?

Survey participation by discipline



Textual analysis methods

- PSU's Literary Informatics Librarian
 - Expertise with textual analysis research
- Voyant Tools - <https://voyant-tools.org/>
 - Web-based tool; Easy and free to use
- Extracted definition survey responses overall and separated out by discipline to upload to Voyant Tools

How do you define data management?



Top Terms

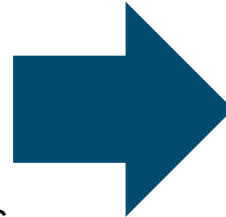
1. Data
2. Storage/Storing
3. Organizing/Organization
4. Management
5. Access

Selected Terms from Data Management Definitions

*“...the compilation of many small practices that make your data easier to **find**, easier to **understand**, less likely to be **lost**, and more likely to be usable during a project or ten years later”* (Briney et al. 2015).

*“... includes data management planning, **documenting** your data, **organizing** your data, improving analysis procedures, **securing** sensitive data properly, having adequate **storage** and backups during a project, taking care of your data after a project, **sharing** data effectively, and finding data for **reuse** in a new project”* (Briney et al. 2015).

*“... is a broad concept that includes processes undertaken to create **organized**, **documented**, **accessible** and **reusable** quality research data”* (Corti et al. 2014).



Find* (find, finding, findable)

Underst* (understand, understood, understanding, understandable)

Los* (lose, loss, lost)

Document* (document(s), documented, documenting, documentation)

Organ* (organize(d), organizing, organization, organise(d), organising, organisation)

Secur* (secure(d), security, securing)

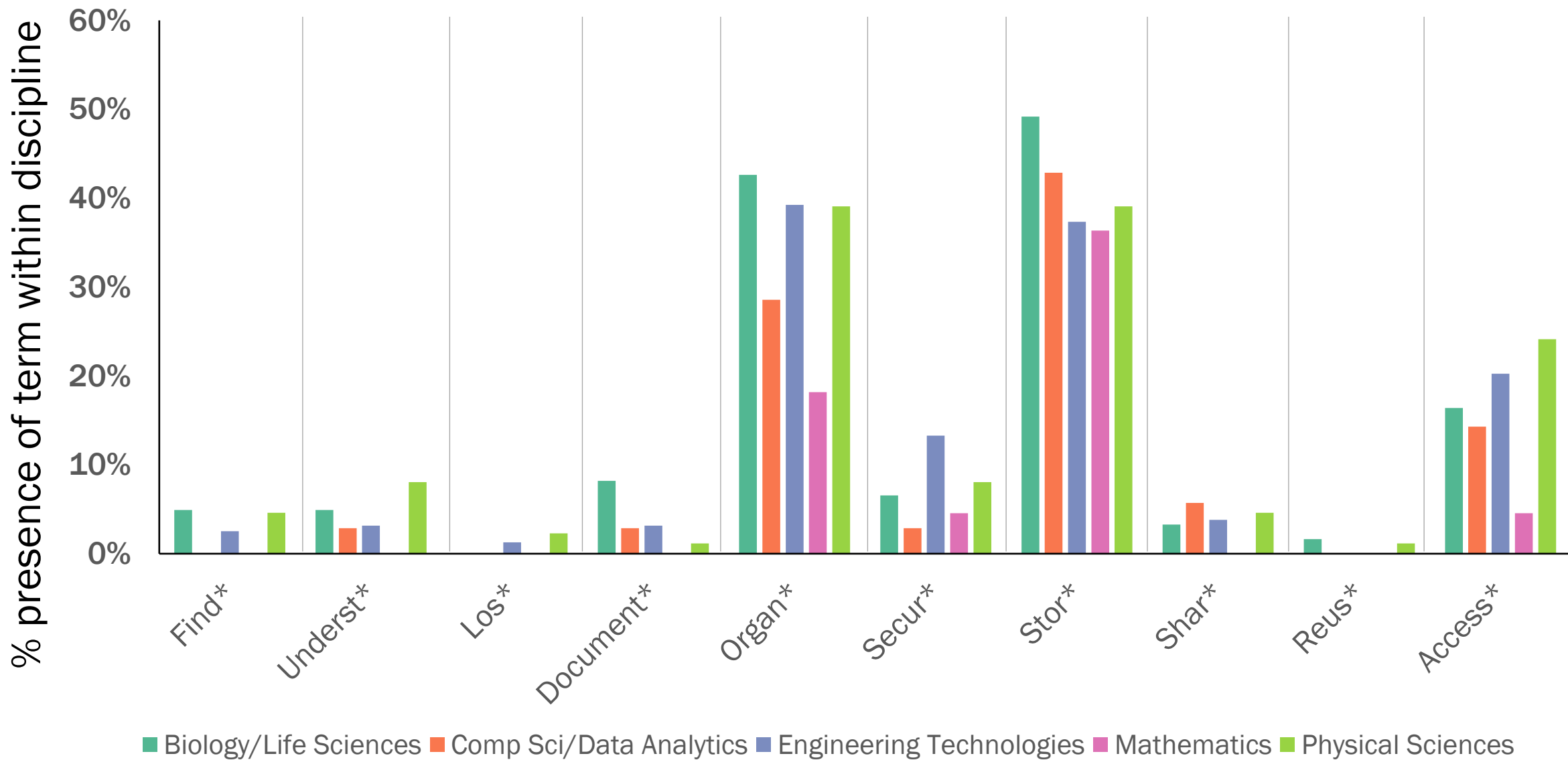
Stor* (storage, store(d), storing)

Shar* (share, sharing)

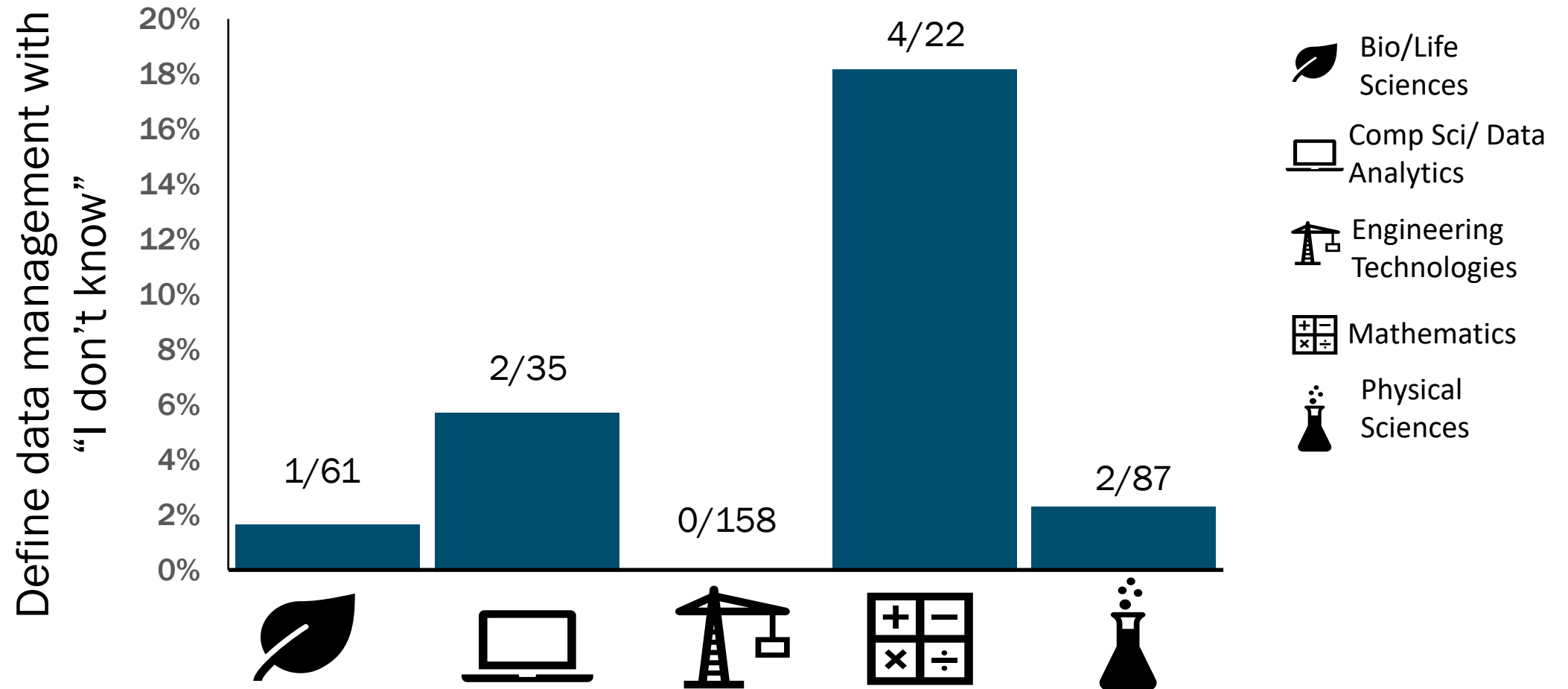
Reus* (reuse, reusable, reusability)

Access* (access(ed), accessibility, accessible, accessing)

How do you define data management?



How do you define data management?



Deeper look...

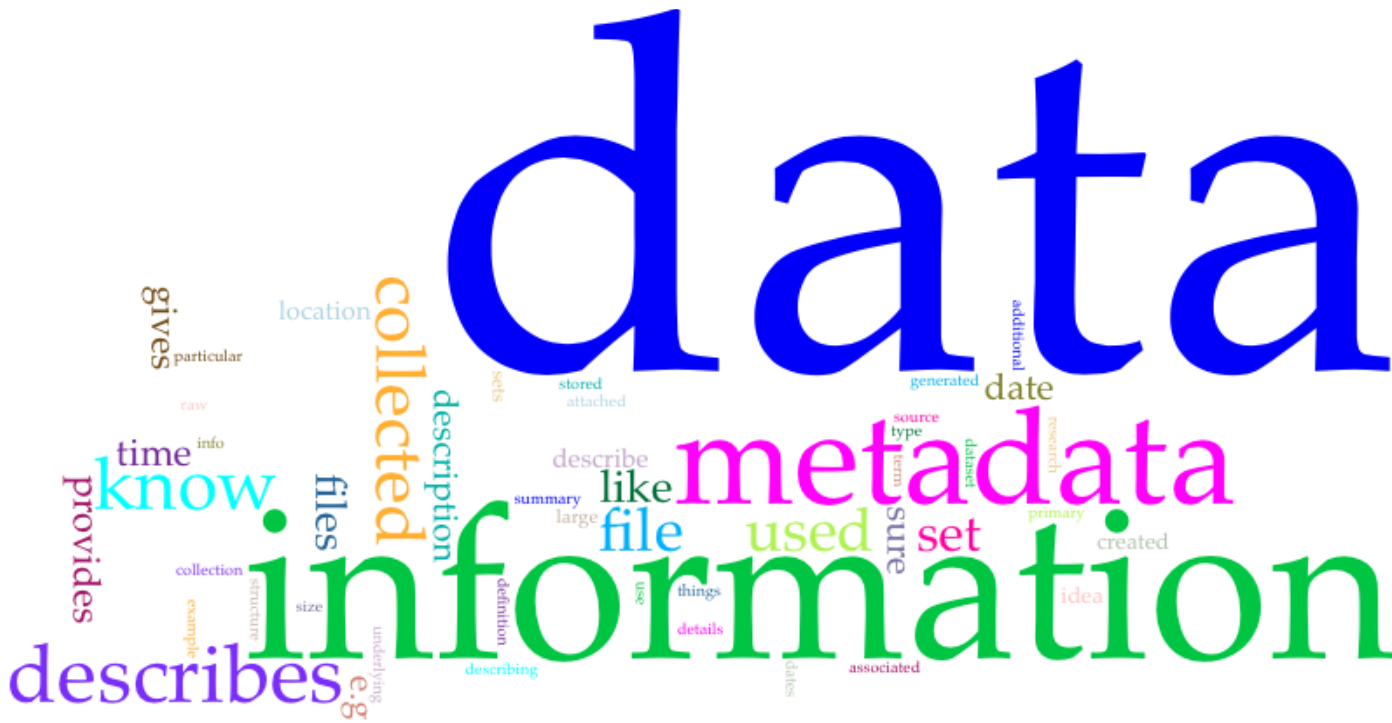
What is left of the word "data"

Contexts		Bubblelines	Correlations	
Document	Left ↑	Term	Right	
DataMan...	Organization and interpretation of all	data	types (observational and quantitative) of	
DataMan...	Context to the left of the keyword. i, and storage of all	data	related to a research project	
DataMan...	that when needed; Uploading all	data	on box/drive; Ways to	
DataMan...	to archived digital and analog	data	through searchable organization hierarchy, naming	
DataMan...	can be used for analysis.	data	management refers to an individual's	
DataMan...	Ability to organize and analyze	data	The process is storing and	
DataMan...	future use. organize and analyze	data	systematically Version control software - Gitlab	
DataMan...	collect, store, transmit, and analyze	data	collected or selected as part	
DataMan...	of organizing, filtering and analyzing	data	Organization protocols for data production	
DataMan...	of organizing, processing and analyzing	data	, and extracting the information from	
DataMan...	ensuring proper security Collecting, analyzing	data	. I dont know I define	
DataMan...	well-organized NA Collection and	data	storage managing data throughout its	

What is right of the word "data"

Contexts		Bubblelines	Correlations	
Document	Left	Term	Right ↑	
DataMan...	The document of the occurrence. results. organization of	data	A good classification of files	
DataMan...	system to use and store	data	Accessing, archiving, and storing data	
DataMan...	How to store and access	data	acquiring, validating, storing, protecting, and	
DataMan...	better store and work with	data	anything that has to do	
DataMan...	Organizing literature sources and your	data	Data management is a method	
DataMan...	and storage and retrieval of	data	Data management is the process	
DataMan...	Storage, documentation, and maintenance of	data	Data storage that is easy	
DataMan...	The organization and storage of	data	Ensuring that data is properly	
DataMan...	storage and and protection of	data	Gathering, storing, analyzing, and processing	
DataMan...	Collection, organization and storage of	data	How you store the data	
DataMan...	how yet. sorting and processing	data	I am not sure I	
DataMan...	of own and peer-published	data	I define it as a	

What are metadata?



Top Terms

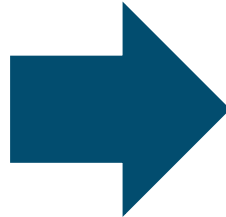
1. Data
2. Information
3. Metadata
4. Describes
5. Collected

Selected Terms from Metadata Definitions

“... is *structured information* that *describes, explains, locates*, or otherwise makes it easier to *retrieve, use*, or manage an information resource. Metadata is often called *data about data* or *information about information*.” (National Information Standards Organization 2004).

“... is *information* about the *context, content, quality, provenance*, and/or *accessibility* of a set of data (University of Wisconsin Madison Research Data Services).

“*Information* about a data set that is *structured* (often in machine readable format) for purposes of search and *retrieval*. Metadata elements may include basic *information* (e.g., title, author, date created, etc.) and/or specific elements inherent to datasets (e.g., spatial coverage, time periods)” (Data Curation Network).



Structur* (structure, structural, structured)
Inform* (informing, information, informational, informative)

Descri* (describe(s), described, describing, description(s), descriptive, descriptors)

Explain* (explain(s), explaining)

Retriev* (retrieve(d), retrieval)

Use* (use(d), reuse, useful) -> exclude user, username

Data about data

Context* (context, contextual)

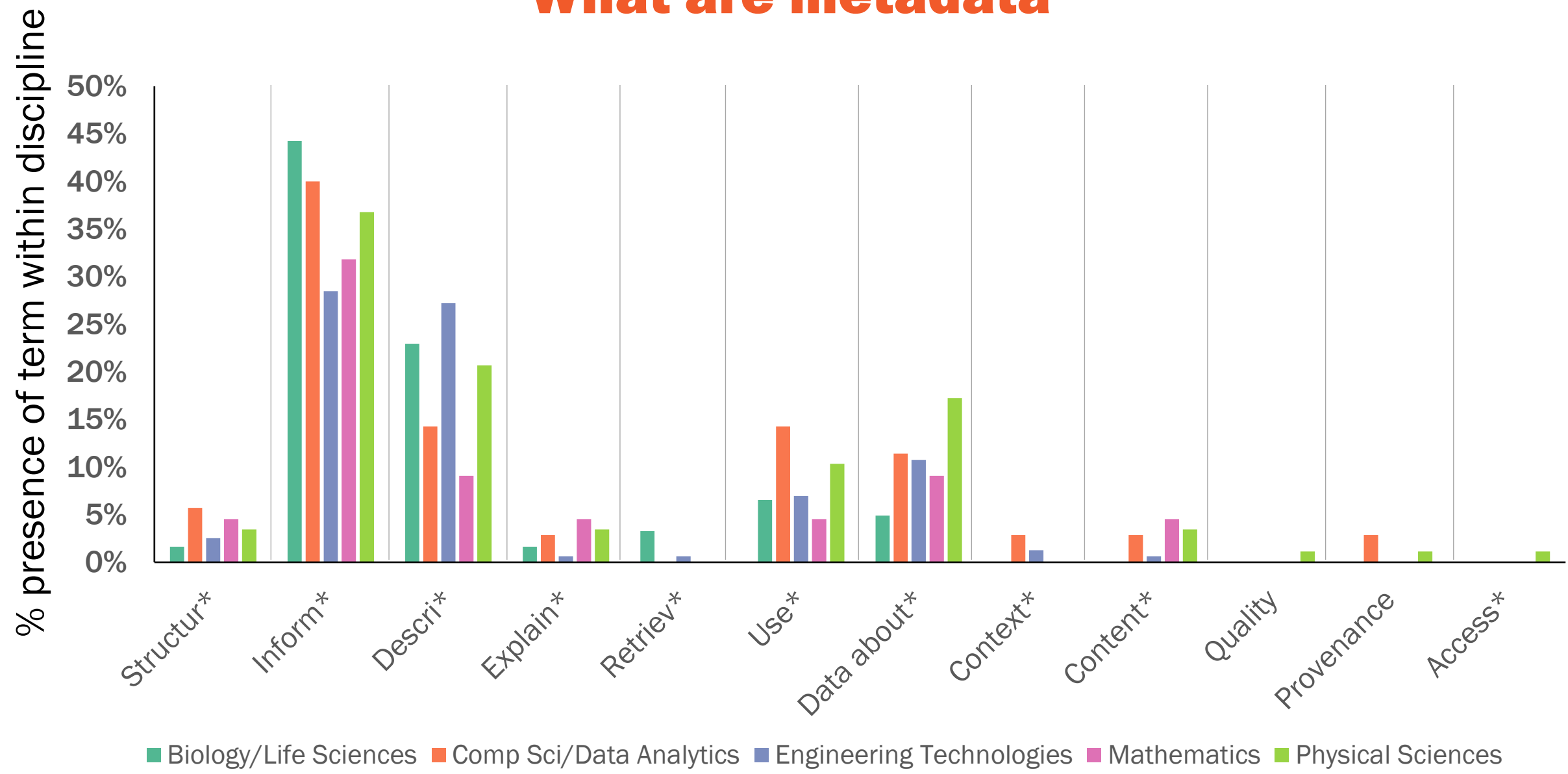
Content* (content(s))

Quality

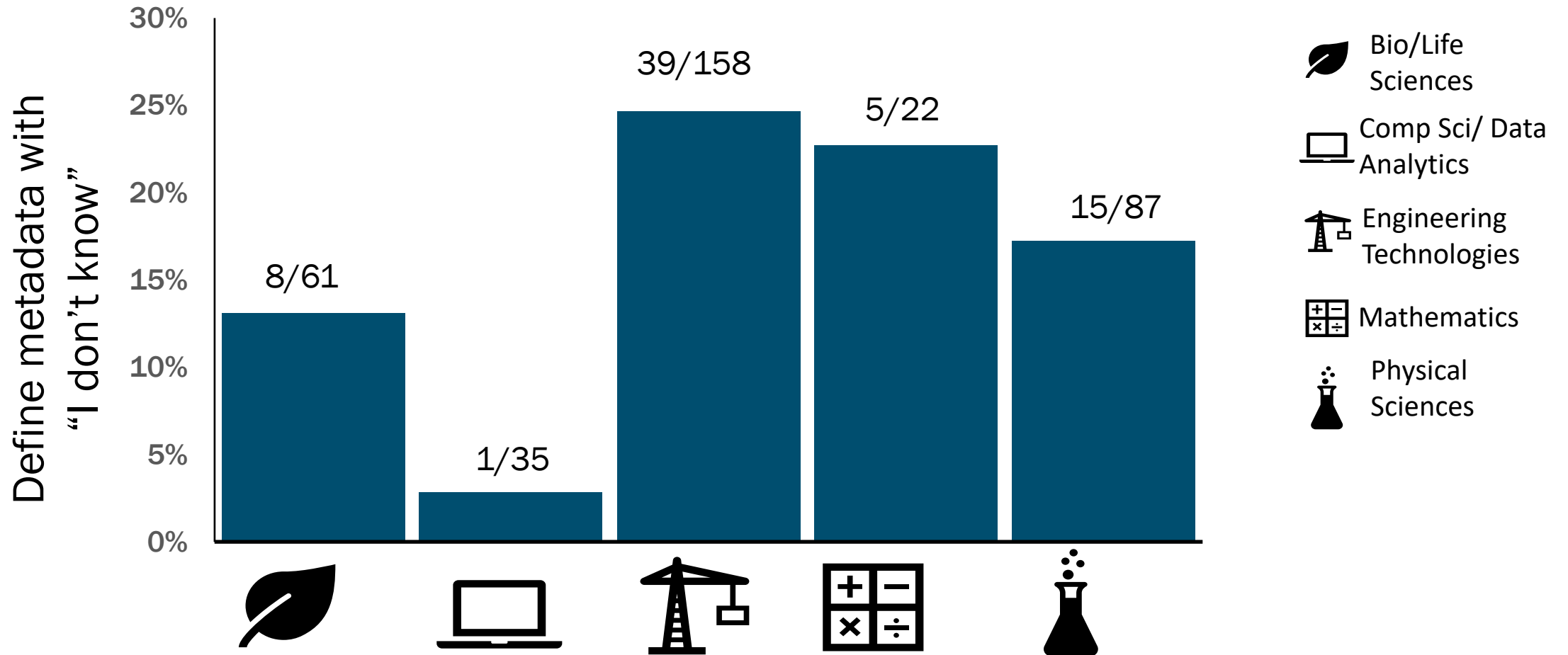
Provenance

Access* (access(ed), accessibility, accessible, accessing) -> exclude access logs

What are metadata



What are metadata?



Deeper look...

What is left of the word "metadata"

Document	Left ↑	Term	Right
WhatAr...	I did not hear about	metadata	Descriptive labels for data. Metadata
WhatAr...	or use. Never heard about	metadata	The data about our resources
WhatAr...	descriptive metadata, structural met...	metadata	,[2] reference metadata, statistical metadata
WhatAr...	data sets to one another.	metadata	are additional information about the
WhatAr...	not sure what metadata are.	metadata	refers to data which depends
WhatAr...	and nature of any artifact.	metadata	are data about the data
WhatAr...	know Data needed for bookkeeping.	metadata	stores additional data that can
WhatAr...	of data to be collected	metadata	is data that provides information
WhatAr...	where/when it was collected	metadata	are the data qualifiers about
WhatAr...	gender, ethnicity, religion are common	metadata	in broad surveys and also
WhatAr...	Sensitive information Data about data	metadata	are pieces of information attached

metadata* × 61 context expand

What is right of the word "metadata"

Document	Left	Term	Right ↑
WhatAr...	The document of the occurrence. I don't know really.	metadata	is defined as the data
WhatAr...	file listings (names and sizes).	metadata	is information about data, such
WhatAr...	for questions about the data.	metadata	is information that pertains to
WhatAr...	definition, but I think that	metadata	is something like data that
WhatAr...	No idea I don't know	metadata	is the additional data associated
WhatAr...	you can call it that.	metadata	is the data summerizing basic
WhatAr...	type. large amounts of data	metadata	is the data within the
WhatAr...	file. Astronomers normally think of	metadata	primarily as all the info
WhatAr...	not sure what metadata are.	metadata	refers to data which depends
WhatAr...	a host of other items.	metadata	should be standardized but it
WhatAr...	assurances In the GIS world,	metadata	standards are defined by the

metadata* × 61 context expand

Preliminary textual analysis

- Top term for both questions is "data" - not very helpful
 - "management" and "metadata" are top terms as well
- Lots of "I don't know", "n/a", "I'm not sure"
 - One (1) "I don't really use data"
- Outreach with researchers needs to account for terminology differences and nuances
 - we need to provide more outreach around terminology

Future textual analysis

- Take deeper look at differences between disciplines by looking at the context of terms
- Analyze new root words
 - Valu* (value, valuable)
 - Preserv* (preserve, preservation)
 - Secur* (secure, security)
- Write and publish related articles (coming soon!)