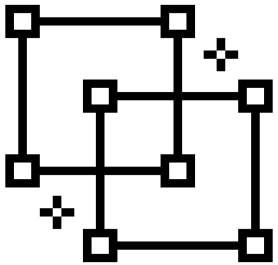
CASE STUDY

Understanding Knowledge Workers' Document Creation, Description, and Storage (CDS) Conventions

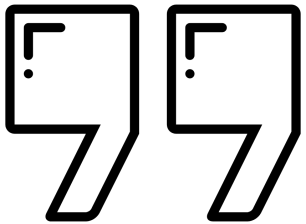
Camille Mathieu
Information Science Specialist
18 June 2019

Jet Propulsion Laboratory
California Institute of Technology

Overview



Study Design



Discussion of Results

Taking Action

Understanding the Problem Space

- Finding information in the enterprise is difficult and costly
- Enterprise information architecture can be highly siloed
- The “Google standard” – user expectations are higher than ever before
- Increasing quantities of data generated daily

A New Approach?

- Industry solutions are diverse: content services, data lakes, machine learning, etc.
- Increasing trend toward software/artificial intelligence solutions, and away from dedicated staff
- JPL is unique - but our findability problem is not

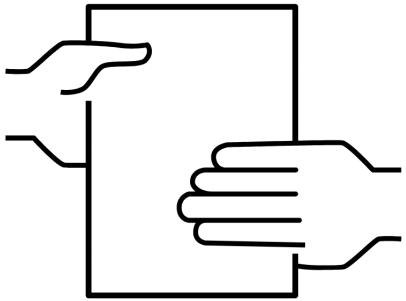
Some Hypotheses

The same features that disadvantage the enterprise can become sources of advantage with process improvement.

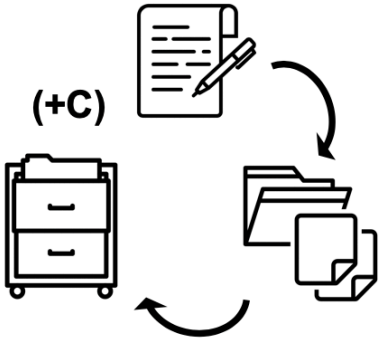
ECM solutions should represent both staff investments as well as software investments.

In the enterprise, it is advantageous to model the knowledge worker not only as the primary content ***consumer***, but also as the primary content ***creator***.

Two Paradigms



Content Creator
vs. Content Consumer



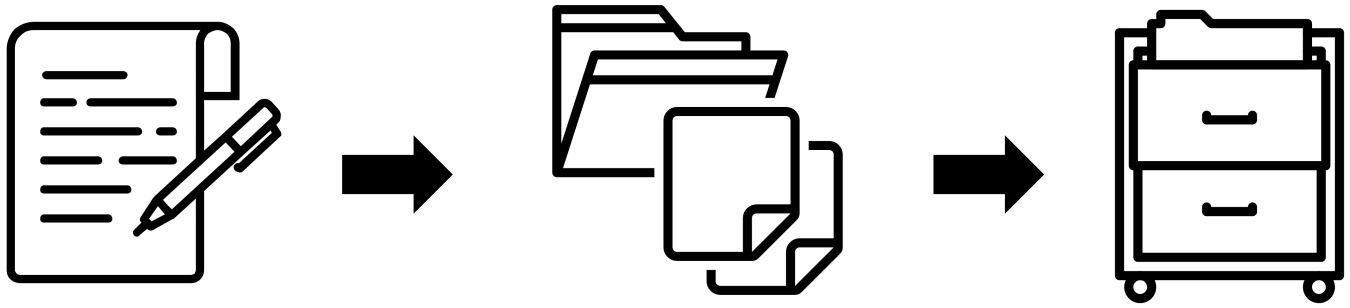
Create/Describe/Store
(+ Consume) Model

Creator vs. Consumer

- What is your relationship to information on the Internet vs. on the intranet?
  - Whereas internet use is mostly passive, intranet users have a more pronounced dual role
- Knowledge workers are both the primary users of enterprise content, and the primary producers
- ECM solutions should take this dual role into account

CDS (+C) Model

How do knowledge workers…



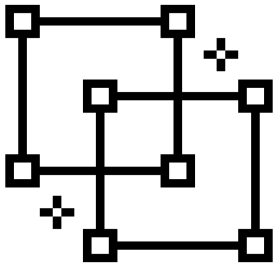
(+ CONSUME)

…enterprise content?

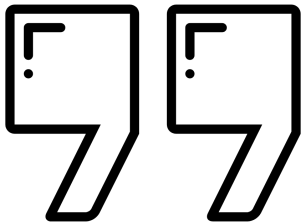
Methods and Timeline

- Study designed during summer 2017 Institute for Research Design in Librarianship (IRDL)
- 35 in-depth interviews conducted February 2018 through October 2018
  - 10 questions asked in 30-45 minute sessions
  - Voluntary response sample set, some convenience sampling
- (Qualitative) Results of interviews normalized and analyzed late 2018 and early 2019

Overview

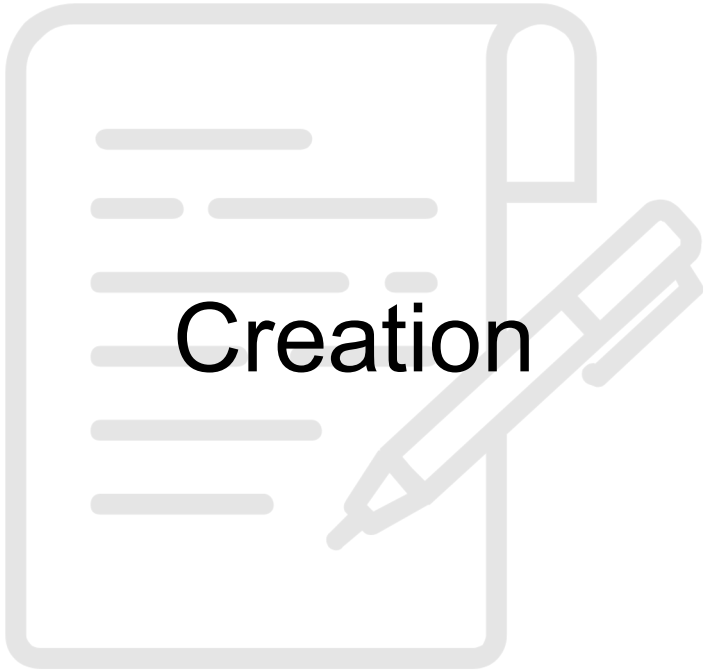


Study Design



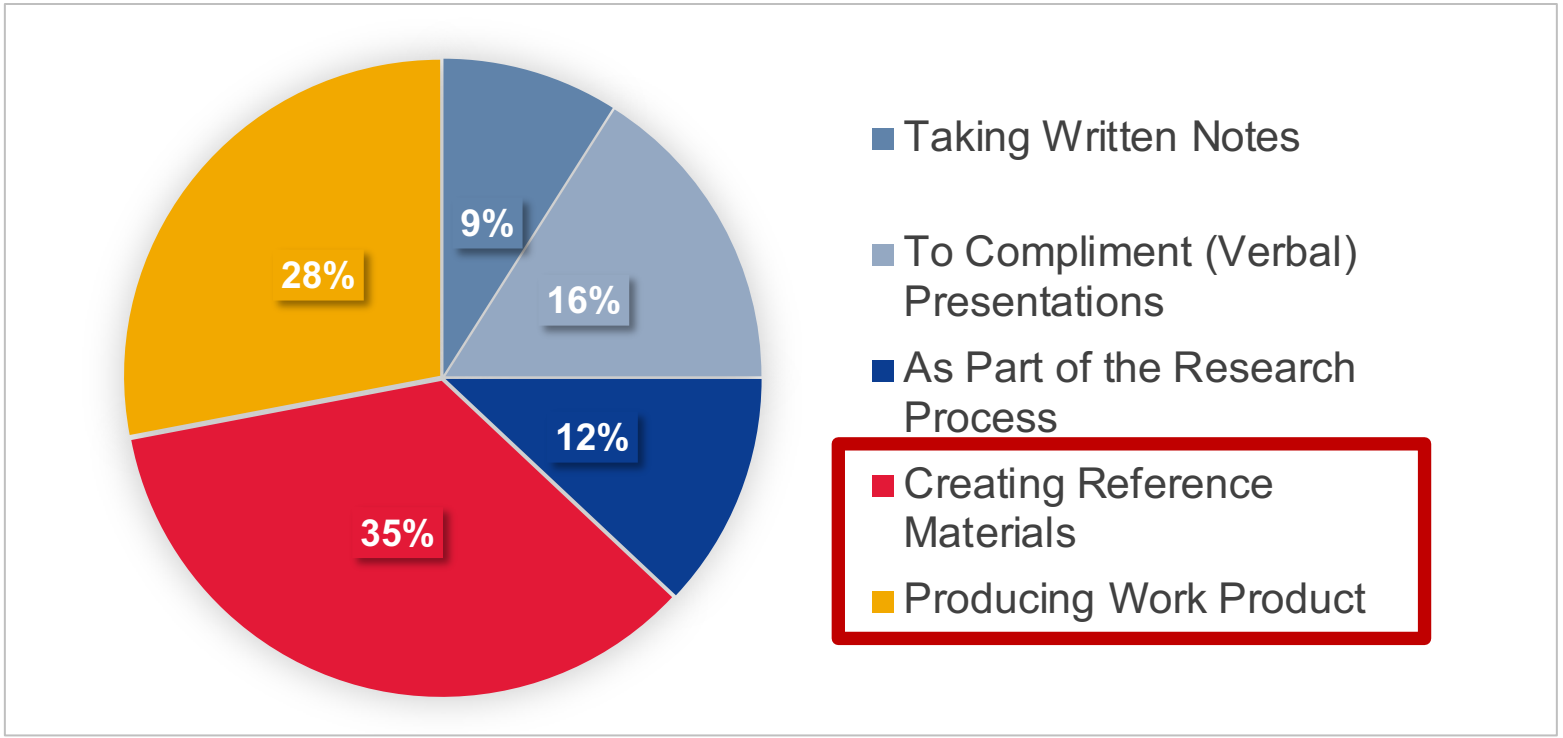
Discussion of Results

Taking Action



Creation

For what reasons do you find yourself creating digital objects at work?



Who is the primary audience of your digital objects?

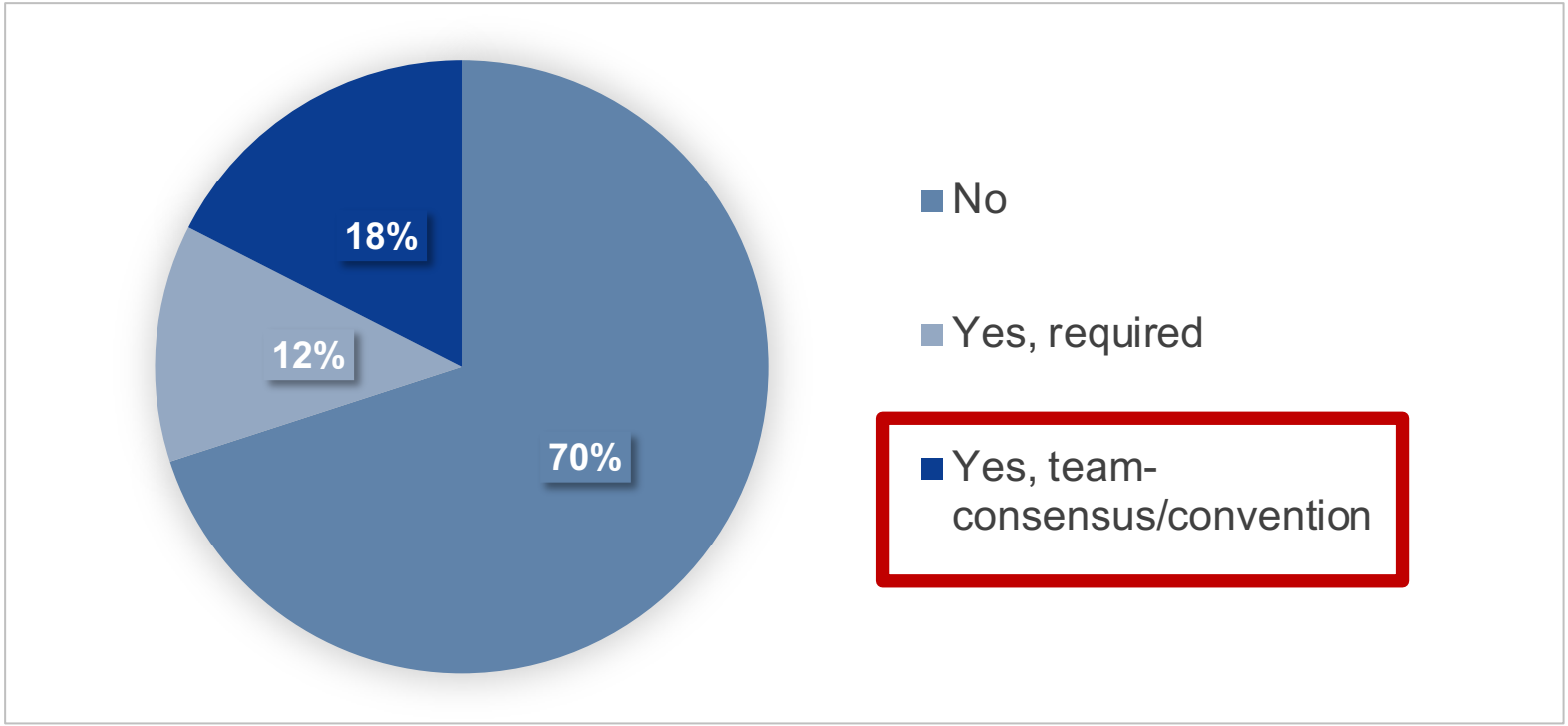
Trends in Creation

- Most knowledge workers are producing either deliverables directly resulting from their work, or documentation on how to do work tasks
- Most of the digital objects generated are intended for use by direct team members and/or managers

Takeaway = Most digital objects represent internal communication about work and processes

Description

Is any descriptive information required for your digital objects?



What kinds of description do you consistently do when creating digital objects?

Organize by folders/folder structure

- Layers indicating subject, date, and status
  - Project/Subsystem
  - Type of Document
  - Draft vs. Final Version
  - Date

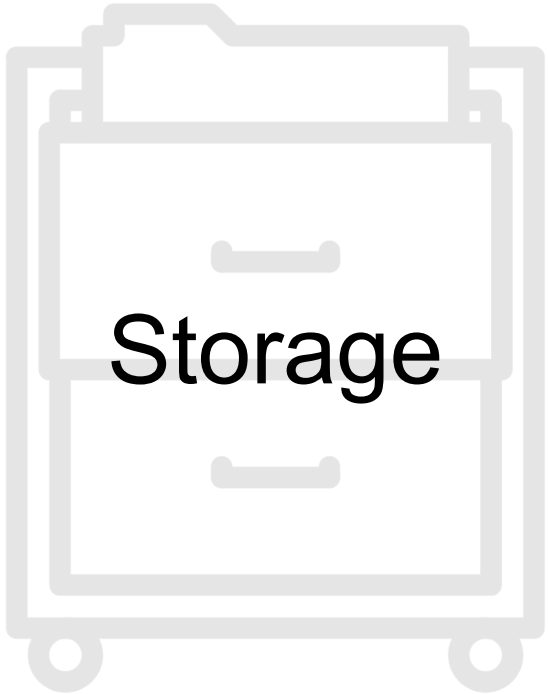
Use file naming convention

- String indicating document title, collaboration, and version
  - Subject/Title
  - Version Number
  - Collaborator initials
  - Project
  - Date

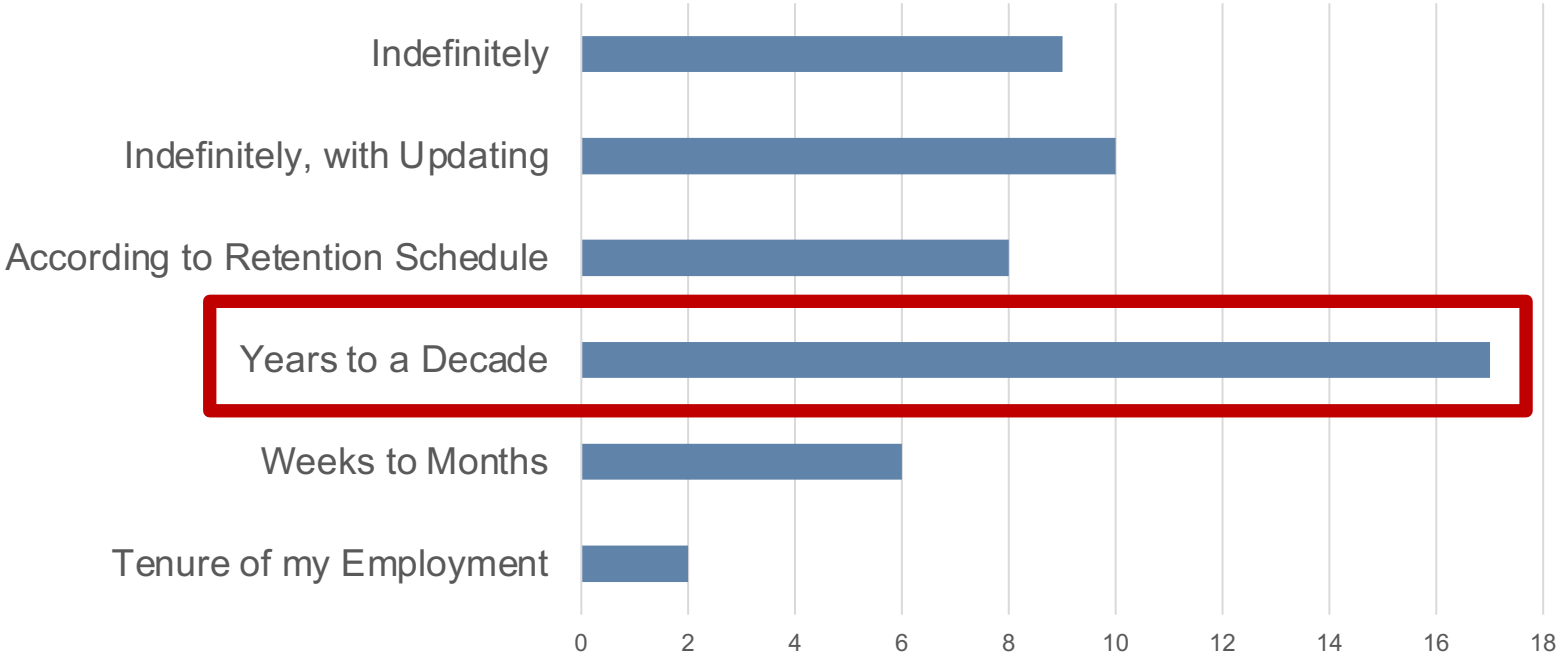
Trends in Description

- Description is rarely mandated
- Teams and individuals are inventing description strategies, usually in the form of folder structures and file naming conventions

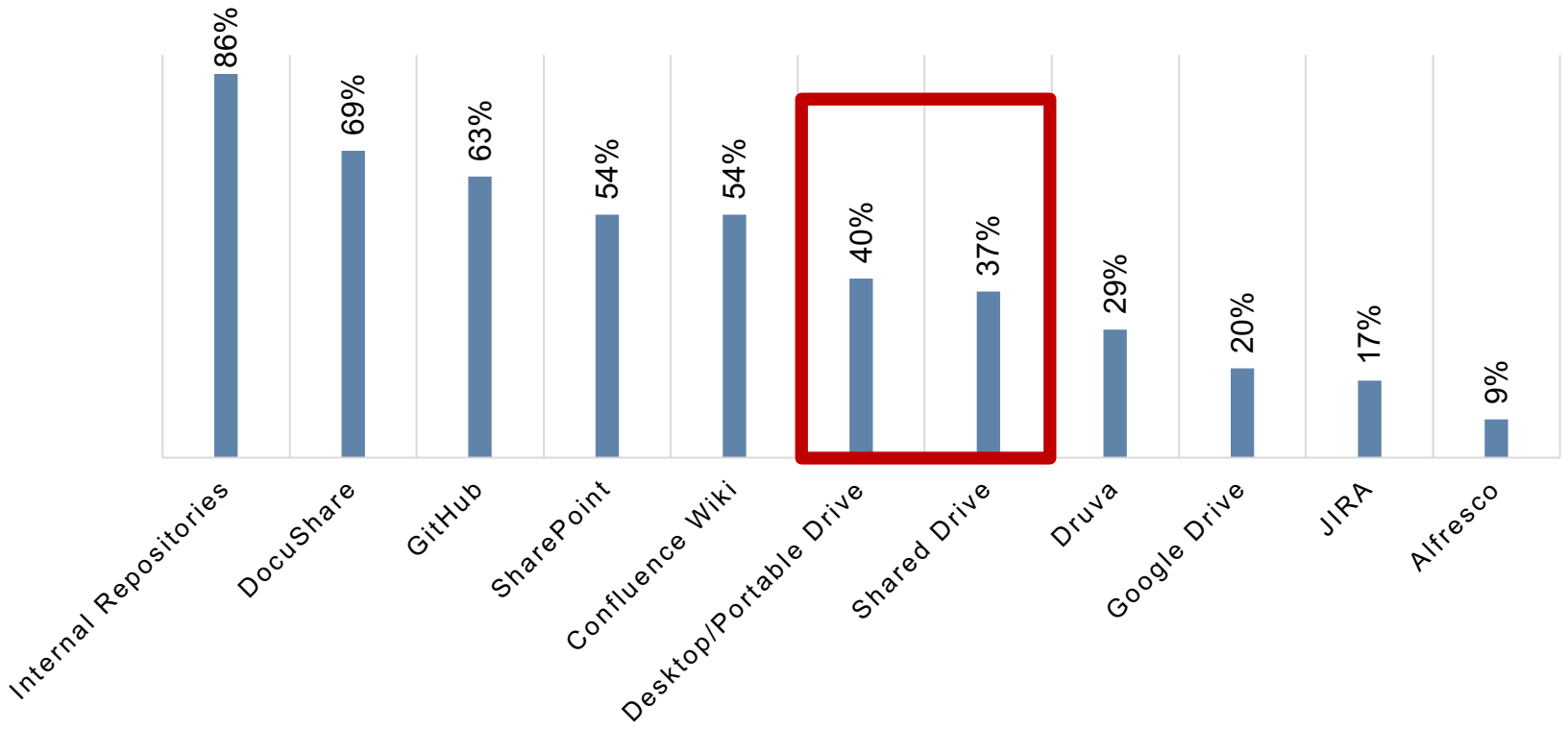
Takeaway = Metadata efforts are mostly ad-hoc and unstructured, typically capturing common fields like date, version, project, and/or contributors


Storage

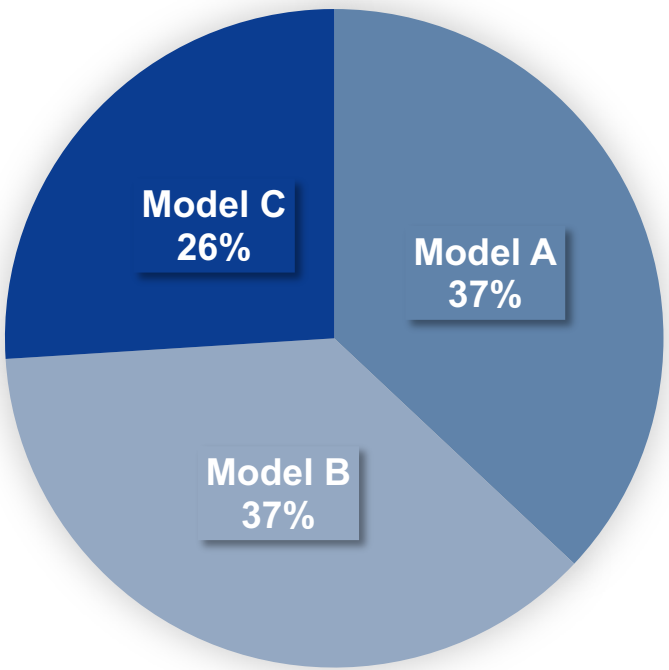
For how long do you expect the digital objects you create to be useful and usable?



Frequency of Repository Use (self-reported)

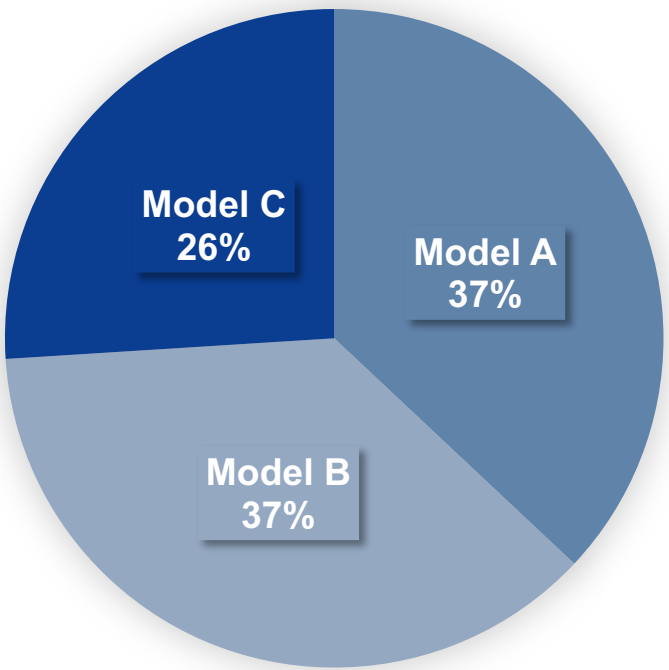


Common CDS Modes by Frequency of Reporting



- A. In-situ creation-description-storage with global access
- B. Created locally, described-stored for global access
- C. In-situ creation-description-storage with local access

CDS Modes as Profiles



Model A (37%)
One-Draft Wonder

Model B (37%)
Pragmatic Publisher

Model C (26%)
Siloed Stewards

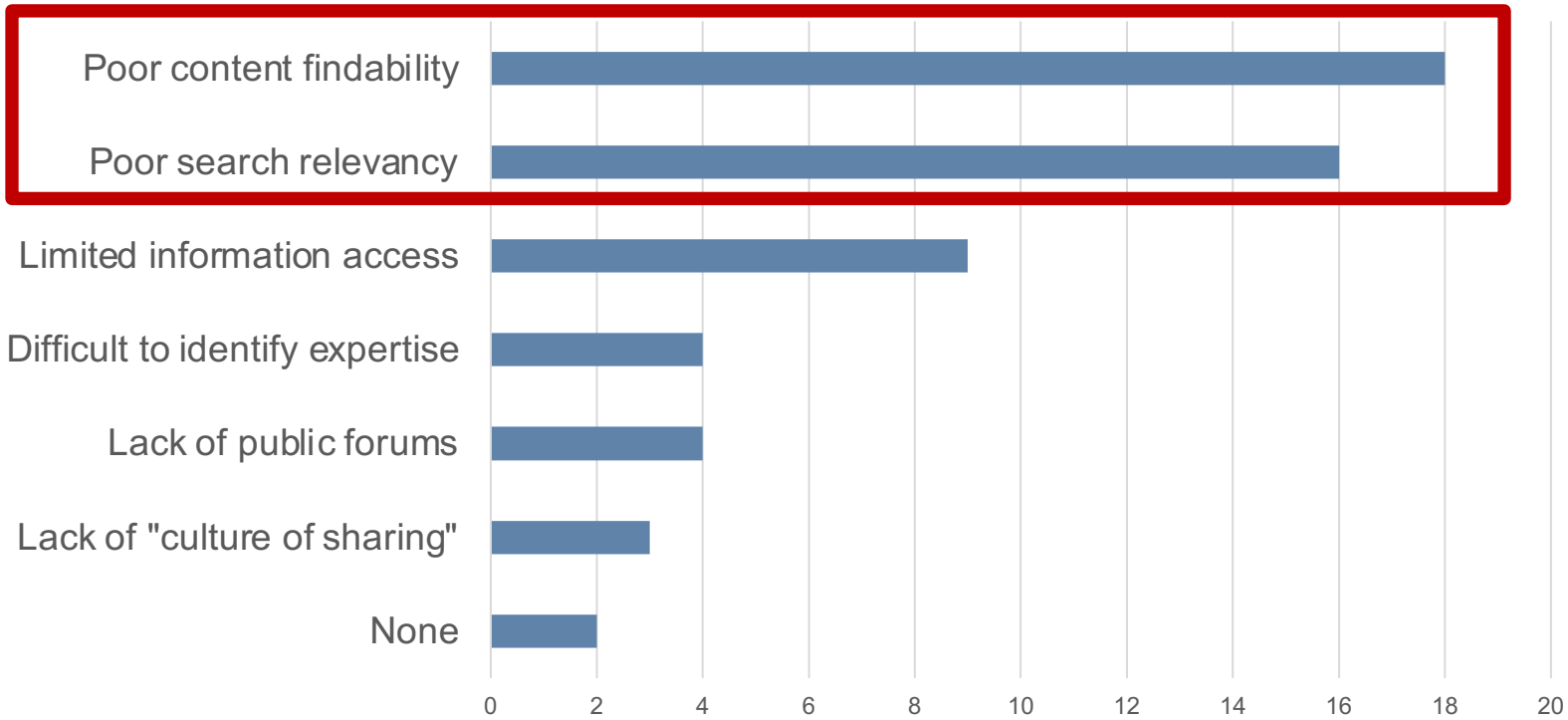
Trends in Storage

- Knowledge workers value their content highly; many believe it should be useful for decades
- About 1/3 of digital objects are stored in siloes accessible (i.e., browsable) only by individuals or immediate teams

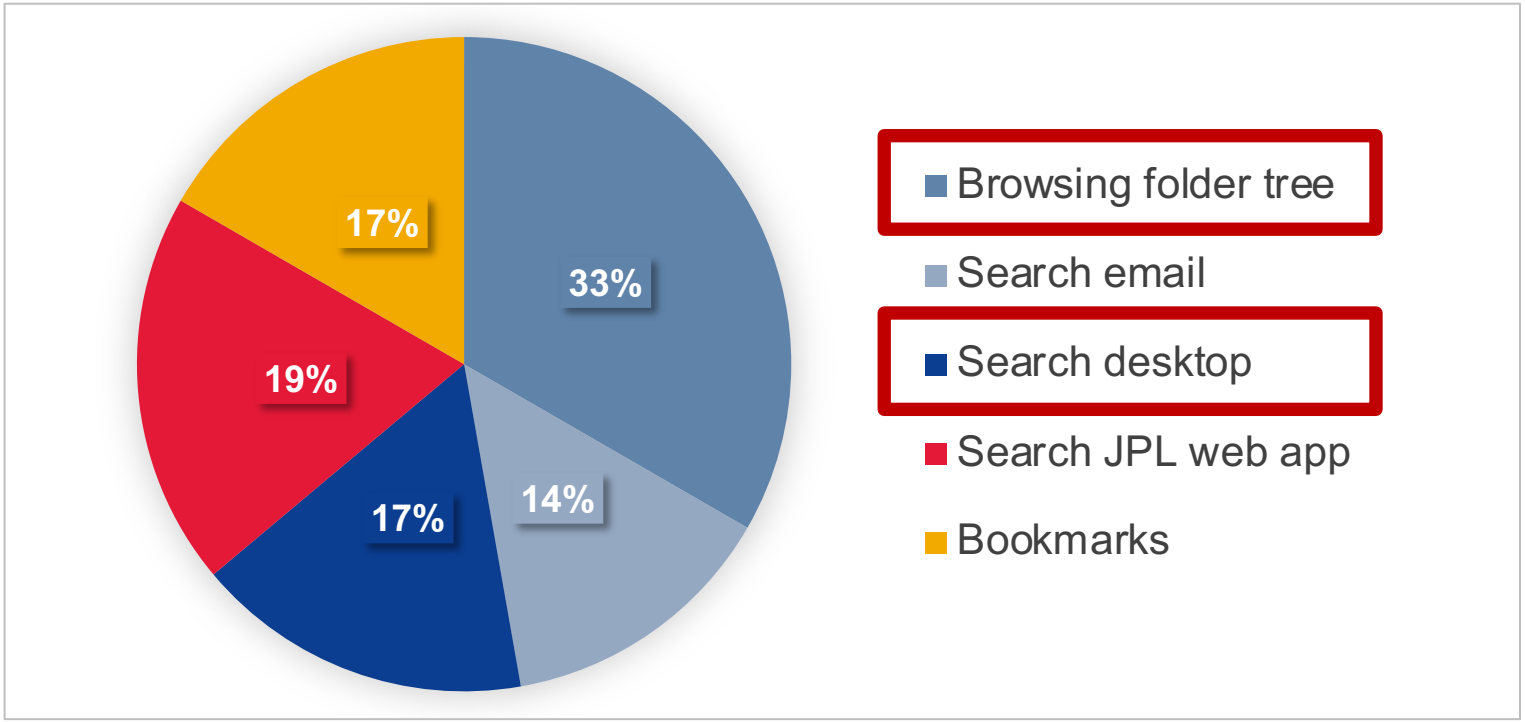
Takeaway = Strategies for storing content do not always reflect the perceived value of the content

(+ Consumption)

Frustrations Using Intranet vs. Internet



How do you refind digital objects that you have previously created, described, and stored?



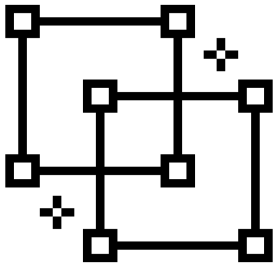
Trends in Consumption

- Most common complaint about intranet is poor search and poor content findability

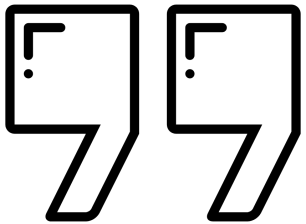
- ~50% of information seeking strategies include searching local files stores and browsing folder trees (consider also that 25-40% of content is stored in local siloes)

Takeaway = There is a gap between search expectations and CDS modes – this gap is exacerbated by reliance on local drives for both finding and storing digital objects

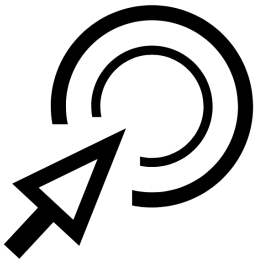
Overview



Study Design



Discussion of Results



Taking Action

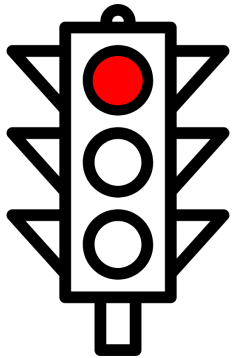
CDS (+C): In Summary

- Most digital objects represent internal communication about work and processes
- Metadata efforts are mostly ad-hoc and unstructured, typically capturing common fields like date, version, project, and/or contributors
- Strategies for storing content do not always reflect the perceived value of the content
- There is a gap between search expectations and CDS practices – this gap is exacerbated by reliance on local drives for both finding and storing digital objects

CDS (+C): Making Improvements

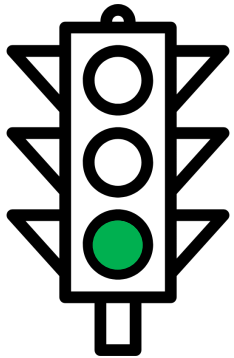
- Creatively optimize the format of information based on the nature of the information being shared; catalog examples of work
- Intervene at points in the CDS process where there is misalignment between knowledge worker practices and intranet goals
- Glean descriptive information from existing folder trees, and allow folder views of content in web apps
- Allow SMEs to define the value and critical context of their content by crowdsourcing taxonomies/folksonomies
- Cultivate open sharing culture with clear guidelines for effectively sharing knowledge – the best search engine in the world cannot find information that is not there

Final Thoughts



Finding information in the enterprise is difficult and costly, and even with advances in technology, the problem remains.

Final Thoughts



Understanding the knowledge worker as both the **creator** and **consumer** of enterprise content, and holistically representing their **CDS (+C) strategies**, can help to design content findability solutions that will support intranet architecture goals throughout the entire document lifecycle.

thank you!

Jet Propulsion Laboratory
California Institute of Technology

Camille Mathieu
cmathieu@jpl.nasa.gov
@searchbrarian
/camillemathieu