

Data Management Plans For Universities and Institutes of Higher Education and Research



Ray Uzwyshyn, Ph.D. MBA MLIS
Director, Collections and Digital Services,
Texas State University Libraries

Data Management Plans: What are They?

- Plans to Manage a Researcher's Data Outputs
- Long Term Archiving/Preservation Strategy
- Ingestion of Various Data Types (from Excel to SPSS to more esoteric forms)
- Permalinking Strategy for Online Citation and Access (DOI: Digital Object Identifier, UNF:University Numerical Fingerprint, Linked Data, Interoperability)
- Paratextual Research Material Archiving/Sharing Possibilities (field notes, docs, data, multimedia and programs)

Types of Data Set Project Repositories

1) Project specific

(usually large single faculty/faculty team projects)

2) Discipline specific

(i.e. Academic Specialization, Purdue Nanohub, Engineering etc.)

3) Institutional Repository

(either institution wide or consortial)



Why are Data Management Plans Required?

Increasingly, major grant agencies require data management plans as mandatory part of the grant proposal process.

(NIH 2003; NSF 2011; NEH, 2013 USDA)



Wordle of the Final NIH Statement on Sharing Research Data, Mandatory 2003

Data Sharing



Currently, 80% of researchers do not share their data

Why Are Data Management Plans Required

- Leverage and make available faculty, departmental and institutional research
- Allow publication of negative data (less research replication)



Wordle of the National Science Foundation's Award and Administration Guide. Chapter VI.D.4, Mandatory 2011

Data Management Plans

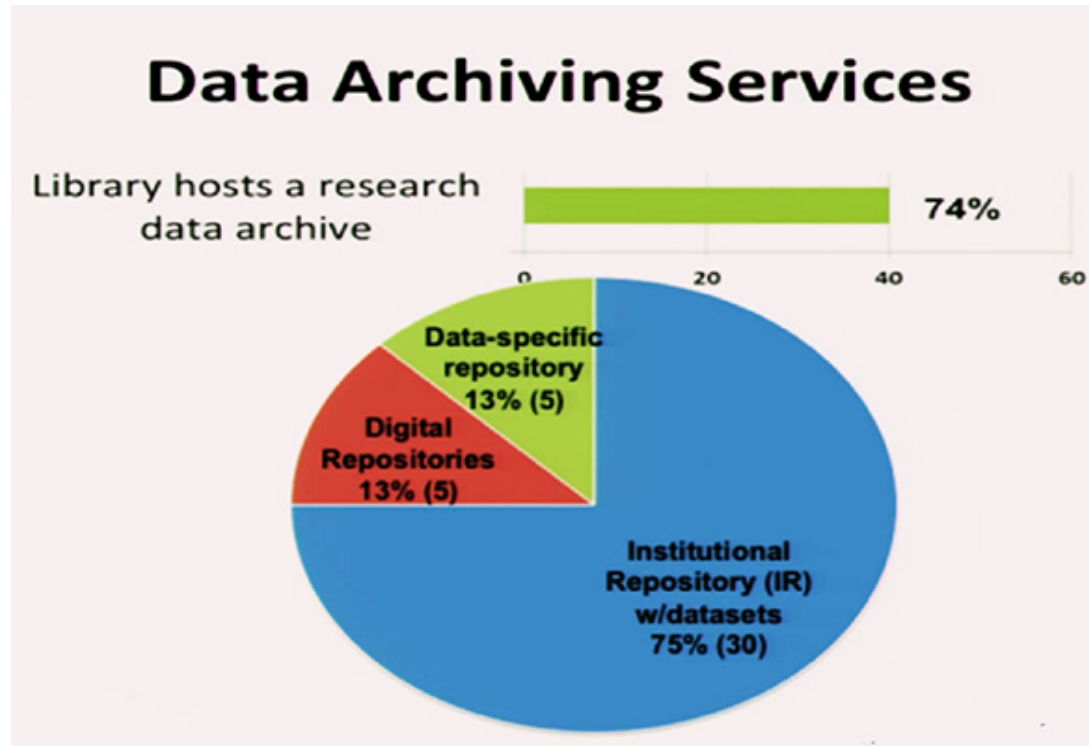
- Part of Evolving Science, Social Science and Humanities Research Process (Accuracy, efficiency, sharing)



Wordle of the data management policy of the Office of Digital Humanities,
National Endowment for the Humanities, 2013

ARL Libraries 2014

Online Data Management Plan Implementation



Fearon, D & Sallans, A. C. (January 2014) Institutional Research Data Management: Policies, Planning, Services and Surveys. Coalition for Networked Information. <https://www.youtube.com/watch?v=rvbrW7S2fes> (54 ARL Libraries currently offer data management services_)

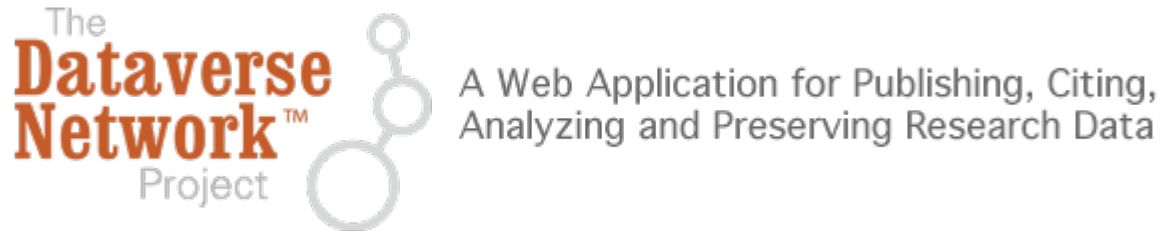
Research Data Management Software

- Online Repositories/Infrastructures created to manage a researcher's data (sharing, archiving, preservation, metadata)
- May be hosted or installed on a university's server
- Each software contains different ranges of management/collaborative options
- Open source and proprietary options

Current DMP Platforms (2014)

Data Archiving Infrastructure	
Primary platform choice	
Inst. Repository w/ Data (top 5)	Data-specific Repository
Dspace	Dataverse
Fedora	Chronopolis
BePress Digital Commons	HubZero (customized)
Hydra	DataConservancy
Drupal	Custom repository

Dataverse: Harvard's Research Data Open Source Solution



Software framework that enables institutions to host research data repositories



Preservation and archival Infrastructure: allows sharing, control, persistent data citation, publishing and management

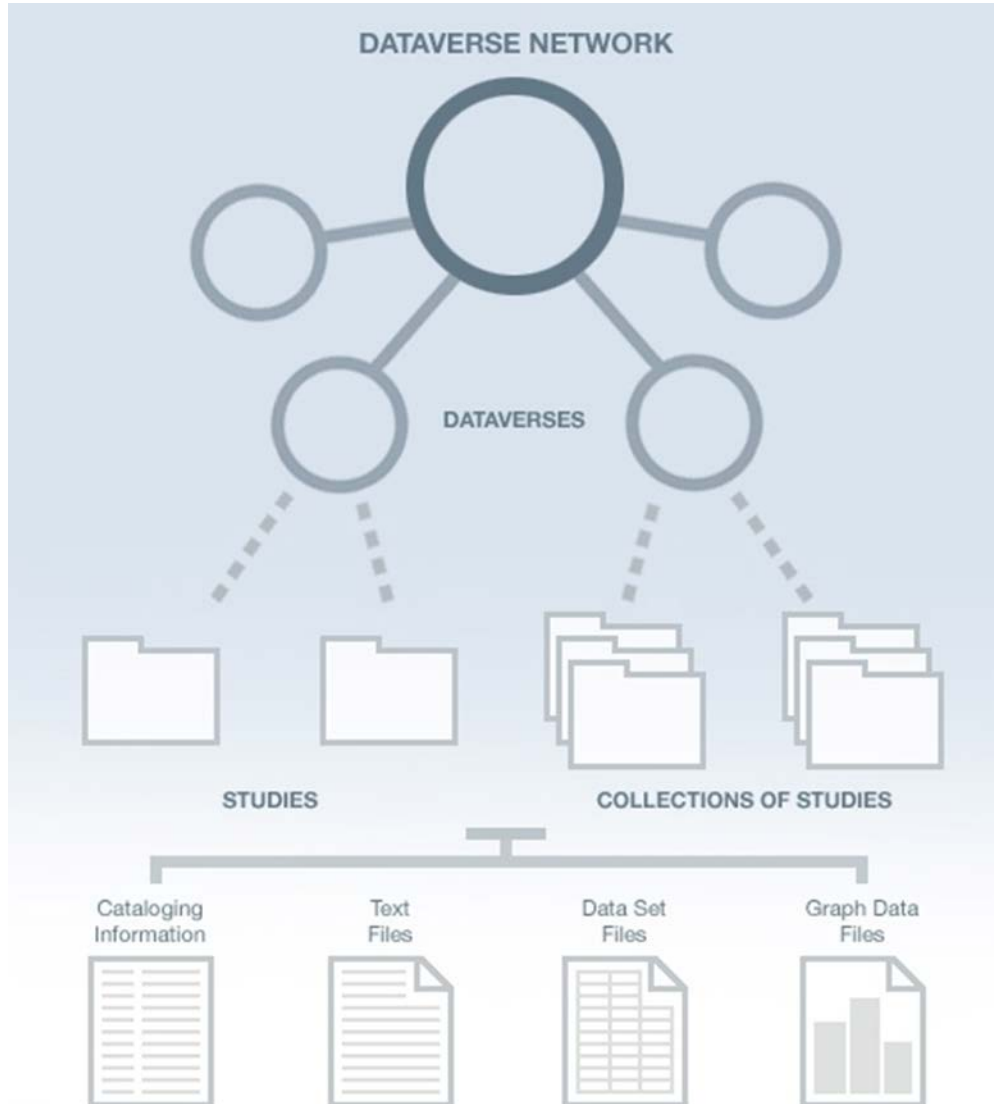
Social Sciences Beginnings (IQSS)

[Data Science](#) (site)

<http://thedata.org>

[Dataverse Open Source Download](#) (Github), [Software Background](#)

Dataverse Network Architecture



[Why the Dataverse Network?](#)
(silent video overview)

[Open Journal Systems
Dataverse Integration](#) (2014)

Research Study Data

Data Set Files

Metadata (Data Describing the data)

Paratextual Research Material
(Methodology, Field Notes etc.)

Graph Data Files

Dataverse Metadata Example

Harvard Dataverse Network

Search   [Create Acc](#)

REPLICATION DATA FOR: A MULTIVARIATE MODEL OF STRATEGIC ASSET ALLOCATION

hdl:1902.1/QBXRSFLBQJUNF:3:ZnYhHkZe2veTJAWaBDpPKA==

Version: 2 – Released: Thu Oct 03 16:46:32 EDT 2013


CATALOGING INFORMATION

Data & Analysis



Comments (0)


Versions

Data Citation

 If you use these data, please add the following citation to your scholarly references. [Why cite?](#)

John Y. Campbell; Yeung L. Chan; and Luis Viceira, 2007, "Replication data for: A Multivariate Model of Strategic Asset Allocation", <http://hdl.handle.net/1902.1/QBXRSFLBQJUNF:3:ZnYhHkZe2veTJAWaBDpPKA==> The Harvard Dataverse Network [Distributor] V2 [Version]

Citation Format  

 Results found in this publication can be replicated using these data.

Original Publication

Campbell, John Y.; Chan, Yeung Lewis; and Viceira, Luis M., 2003, "A multivariate model of strategic asset allocation," Journal of Financial Economics, Elsevier, vol. 67(1), pages 41-80: [article available here](#)

Publications

John Y. Campbell & Yeung Lewis Chan & Luis M. Viceira, 2001. "A Multivariate Model of Strategic Asset Allocation," NBER Working Paper National Bureau of Economic Research, Inc. [article available here](#)

Campbell, John Y & Chan, Yeung Lewis & Viceira, Luis M, 2001. "A Multivariate Model of Strategic Asset Allocation," CEPR Discussion Paper 3070, C.E.P.R. Discussion Papers. [article available here](#)

Data Citation Details

Title	Replication data for: A Multivariate Model of Strategic Asset Allocation
Study Global ID	hdl:1902.1/QBXRSFLBQJ
Authors	John Y. Campbell (Harvard University); Yeung L. Chan; and Luis Viceira
Producer	 John Y. Campbell HARVARD Faculty of Arts and Sciences DEPARTMENT OF ECONOMICS
Production Date	2003
Funding Agency	National Science Foundation; Hong Kong RGC Competitive Earmarked Research Grant (HKUST 6965/01H); Division of Research of the Business School

Dataverse Metadata Example

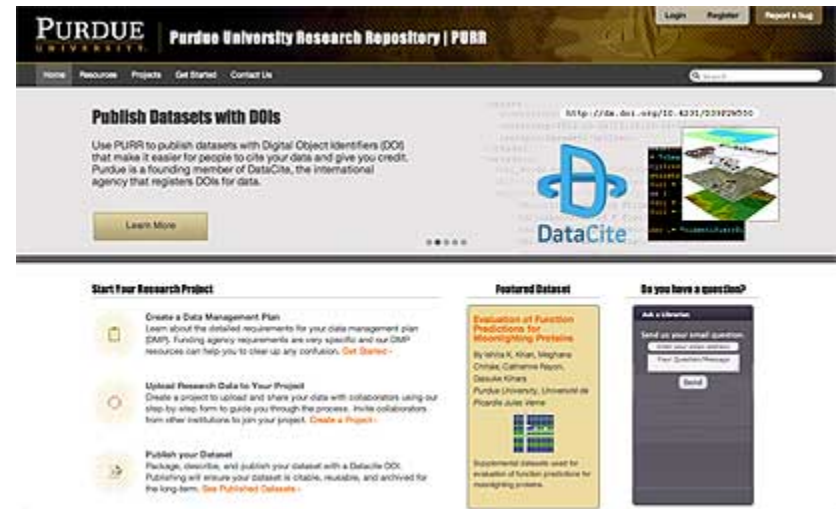
Files			Metadata	Versions
Citation Metadata ^				
Title		Data from "Social determinants of unmet hospitalisation need amongst the poor in Andhra Pradesh, India: A cross-sectional study."		
Author	Name	Affiliation		
	Identifier	Identifier Scheme		
Description		The dataset is of a health survey amongst the 21.5 million poor families of the Indian state of Andhra Pradesh conducted during April and May 2013. The dataset captures individual characteristics and household characteristics of the past 365 days and was used to analyse the unmet hospitalisation need in the Indian State of Andhra Pradesh. Data was collected by 2022 trained field staff of Aarogyasri Health Care Trust (AHCT) of Government of Andhra Pradesh using a questionnaire modelled after that used for the health surveys by National Sample Survey Organisation of India.		
Subject		Medicine, Health & Life Sciences		
Keyword		unmet hospitalisation need		
Production Date		2013-06-01		
Depositor		Privileged, admin		
Deposit Date		2013-08-03		

PURR and Hubzero: Purdue's Data Management System

- 1.) Create Data Management Plans
- 2) Collaborate with other Researchers
- 3) Publish Data Sets (Purdue can publish a DOI: Digital Object Identifier for Data Sets)
- 4) Archive Data Sets

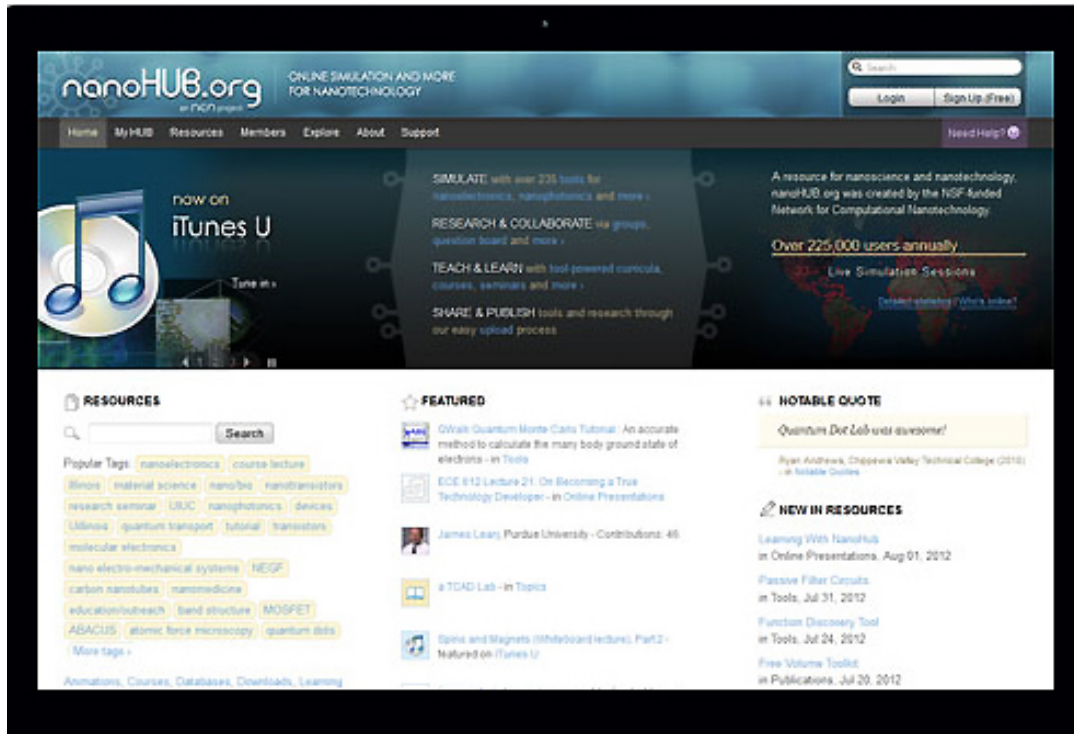
Boilerplate text for data management proposals available

Purr is part of Hubzero platform for scientific collaboration (Originally Nanohub)



- [Purr: Purdue University Research Repository](#) (video)
- [Purr Site \(Proprietary to University\)](#)
- [Purr Background](#)

Hubzero: Open Source Platform for Scientific Collaboration



Research Collaboration and Data Management Solution

Research Data Types

Spreadsheets

Instrument or Sensor Readings

Software Source Code

Surveys

Interview Transcripts

Images and Audiovisual Files

- <https://hubzero.org/>
- [Getting Started](#), [Downloadable](#) and [Hosted Options](#)
- [Hubzero Video](#), [Hubzero2](#)

Hubzero/Purr Customization

Start Your Research Project



Create a Data Management Plan

Learn about the detailed requirements for your data management plan (DMP). Funding agency requirements are very specific and our DMP resources can help you to clear up any confusion. [Get Started ›](#)



Upload Research Data to Your Project

Create a project to upload and share your data with collaborators using our step-by-step form to guide you through the process. Invite collaborators from other institutions to join your project. [Create a Project ›](#)



Publish your Dataset

Package, describe, and publish your dataset with a Datacite DOI. Publishing will ensure your dataset is citable, reusable, and archived for the long-term. [See Published Datasets ›](#)

Hubzero/Purr

File Level Research Data View

Localization of potential regulatory signals in the human genome

By Minou Bina, Phillip J Wyss

Purdue University

Supplementary materials for the publication entitled "Discovering sequences with potential regulatory characteristics." Bina M, Wyss P, Lazarus SA, Shah SR, Ren W, Szpankowski W, Crawford GE, Park SP, Song XC. Genomics 93: 314-22 (2009)

Listed in [Datasets](#)

About

Supporting Docs

Versions

Reviews

Questions

Citations

Abstract

In the human genome, the expression of protein-coding genes is in part regulated by specific DNA sequence elements known as proximal and basal promoters. These elements are relatively short and usually correspond to transcript start sites (TSSs) determining the genomic DNA sequences that control the expression of human genes. However, it has been difficult to identify these regions via various experimental strategies including chromatin immunoprecipitation assays (ChIPs). To develop a complete catalog of promoter regions would provide a rich source of data for designing tools to predict the position of regulatory signals. We designed a scheme to create density plots of sequences that appear frequently upstream of TSSs.

Figshare/Cloud based/Proprietary



Repository where users make their research available in citable, shareable and discoverable manner

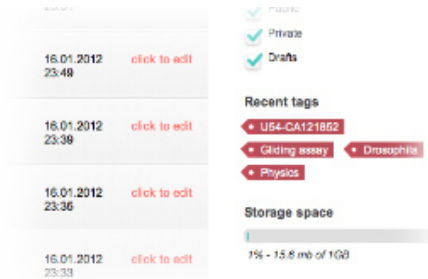
Figures, datasets, media, papers, posters presentations and file sets can be disseminated In a way that the current scholarly publishing Model does not allow

Open Source Platform for Sharing Research

[Figshare](#) (video)

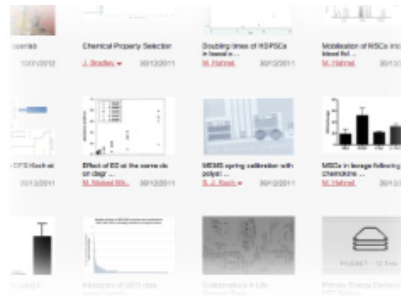
[Figshare for Institutions](#) (Video)

Figshare Features (Cloud Based/Proprietary)



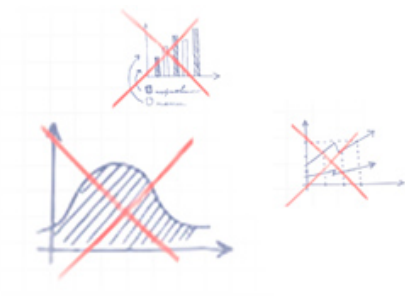
1GB of private space

taggable and easily filtered, your research data
is better managed and easy to locate



Unlimited public space

upload to your heart's content
the more - the better

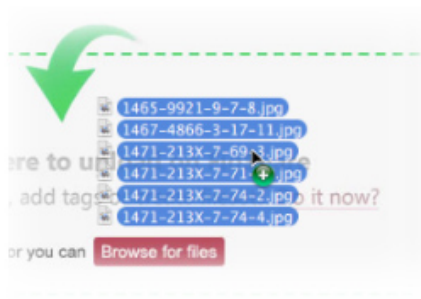


Publish negative data

all published research is citable



Upload all formats



Quick & simple upload



Cloud based

One Size Does Not Fit All Data Project Needs

Types of Data Projects (Sizes)

1) Normal range

Files/Data Fit on Server/Cloud, may be uploaded, Dataverse, Purr)

2) Large Projects

(Data may require specialized university IT Support, i.e. terabyte/petabyte tape drives)

3) Very Large

(Projects require consortial possibilities, national models, Chronopolis, Texas Advanced Computer Center, DEEPN, Duracloud)






DMP Documentation/Policy Tool



[Overview Video](#)

Customizable
Plan Outline Tool
Resource Links
Supports All
Major Funders

 PUBLIC DMPs	 DMPTOOL NEWS	 DMPTOOL HELP
List of sample data management plans provided by DMPTool users. » CAREER: Parietal Cortex and the Transformation of Spatial Cognition into Action	Latest information about data management and the DMPTool. » US Dept of Energy data management requi... » Medical Research Service	Overview of how to use the tool, plus resources and guidance on data management. » Frequently Asked Questions » Create a DMP

<https://dmptool.org/>
California Digital Library

Further Links/References

- ARL NSF Data Sharing Policy and Resource Links: <http://www.arl.org/focus-areas/e-research/data-access-management-and-sharing/nsf-data-sharing-policy#.VAXZ3E1OWUk>
- California Digital Library DMT Tool: <https://dmptool.org/>
- Dataverse. <http://thedata.org/>
- Dataverse (Data Science Site). <http://datascience.iq.harvard.edu/dataverse>
- Purr. (Purdue Institutional Data Repository). <https://purr.purdue.edu/>
- Hubzero. <https://hubzero.org/>
- Figshare. <http://figshare.com/>
- Research Data Management. Principles, Practices, and Prospects (November 2013). *Council on Library and Information Resources*. <http://www.clir.org/pubs/reports/pub160>
- Fearon, D & Sallans, A. C. (January 2014). Institutional Research Data Management: Policies, Planning, Services and Surveys. Coalition for Networked Information. <https://www.youtube.com/watch?v=rvbrW7S2fes> (video presentation)
- Data Management for Libraries: (LITA Guide) <http://www.alastore.ala.org/detail.aspx?ID=10737>
- *NMC Horizon Report: 2014 Library Edition*. <http://cdn.nmc.org/media/2014-nmc-horizon-report-library-EN.pdf>
- “Research Data Management”. pp. 6-7 and pp 24 – 45.
- Holden, J. Memorandum for Heads of Executive Departments and Agencies: Increasing Access to the Results of Federally Funded Research (2013). http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
- Research Data Management in the Arts and Humanities (2013). University of Oxford. <http://www.dcc.ac.uk/events/research-data-management-forum-rdmf/rdmf10-research-data-management-arts-and-humanities> (Conference Presentations)

Comments/Questions

