PDS
Planetary Data System

PDS Roadmap and Discussion (Update on HQ Briefing)

Ralph L. McNutt, Jr.[1]

[1]The John Hopkins University Applied Physics Laboratory

Wednesday, 19 April 2017
1:15 PM – 3:15 PM

San Rafael Room, The Westin Hotel
Pasadena, CA

Roadmap Study Status

- **Working to finalize a *CONSENSUS* Roadmap Study Report**
- **Aim point remains end of this month to finalize the wording**
- **Briefed at NASA HQ on 4 April – David Schurr and others in attendance**
  - David brought up several items
  - Suggested that funds might be available to deal with some of these

Issues Being Worked

- **The Roadmap is a written document in terms of "Findings" and "Remediations"**
  - The RST is not a FACA committee
  - ~40 page report with ~60 pages of Appendices
- **Most of the document is complete**
- **Content and supporting language still being worked on 4 potential items – all interrelated**
  - 3 Findings: "Node Structure," "Governance," and "Transparency"
  - "Executive Summary"

PDS
Planetary Data System

NASA Planetary Data System (PDS) Roadmap Progress Report

Ralph L. McNutt, Jr.[1], Ross A. Beyer[2,3], M. Katherine Crombie[4], Denton Ebel[5], Lisa R. Gaddis[6], Amitabha Ghosh[7], Emily S. Law[8], Thomas H. Morgan[9], Flora Paganelli[3,10], Anne C. Raugh[11], Matthew S. Tiscareno[3], Thomas C. Stein[12]

[1]The John Hopkins University Applied Physics Laboratory, [2]NASA Ames Research Center, [3]SETI Institute, [4]Indigo Information Services, LLC, [5]American Museum of Natural History, [6]Astrogeology Science Center U. S. Geological Survey, [7]Tharsis Inc., [8]Jet Propulsion Laboratory, [9]Goddard Space Flight Center, [10]American Public University System, [11] University of Maryland, [12]Washington University

Tuesday, 4 April 2017
2:00 PM – 3:00 PM

NASA Headquarters MIC 3B (3J42 – B)
Washington, D.C.

Status

- Roadmap Study Team (RST) is up and running
- Telecoms now held every week (except for LPSC)
- Face-to-Face meetings held 13-14 October 2016 and 8-10 February 2017 in Towson, Maryland
- Current Draft report is 44 pages plus 57 pages of Appendices
- Poster on progress presented at Lunar and Planetary Science Conference
- Target completion of written report for April 2017 – prior to next PPBE – and still on track

The Planetary Data System aka PDS

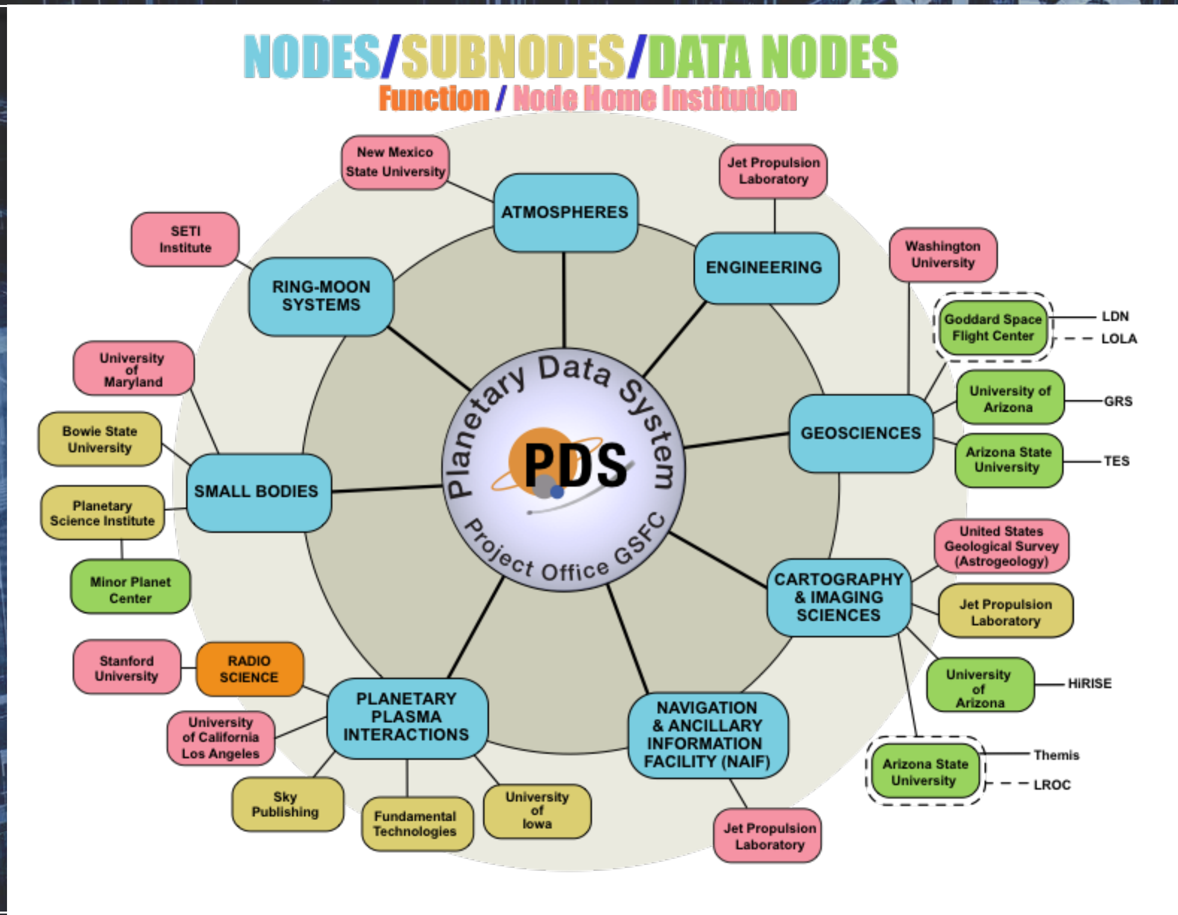
- Up and running since 1990
- Responded to concerns raised that planetary data being returned by scientific satellites was in danger of being lost
- The Committee on Data Management and Computation (CODMAC) was chartered by the National Academies to address the issue
- On the basis of peer-reviewed proposals, discipline-oriented nodes were selected to form the core of the PDS.

PDS Functionality

- Archives electronic data products from NASA planetary missions, sponsored by NASA's Science Mission Directorate
- Makes peer-reviewed, documented, data available online to scientists and to the public without charge
- Works closely with project teams to help them design well-engineered products that can be released quickly
- Works closely with the community to provide higher order data products (and new archive materials) by supporting investigators in NASA's Planetary Data Archiving, Restoration, and Tools Program (PDART).
- Provides teams of scientists to help users select and understand the data and offers special processing for products tailored to the needs of individual users.

Current Organizational Structure

- Federated system
- Eight *Discipline Nodes* (competed and distributed)
- Two technical support nodes (JPL)
- Project Office (GSFC)



PDS Requirements

1) PDS will provide expertise to guide and assist missions, programs, and individuals to organize and document digital data supporting NASA's goals in planetary science and solar system exploration.

2) PDS will collect suitably organized and well-documented data into archives that are peer reviewed and maintained by members of the scientific community.

3) PDS will make these data accessible to users seeking to achieve NASA's goals for exploration and science.

4) PDS will ensure the long-term preservation of the data and maintain their usability.

What Is PDS *Not* Intended to Do?

- PDS serves complex, high-volume data from a wide variety of instruments to a large number of stakeholders spread across the globe
- While PDS responsibilities are extensive, it is important to note that per NASA policy and practice there are numerous activities that are outside the purview of PDS

What PDS does not do (1)

- PDS does not set the mission archiving requirements.
- PDS does not develop requirements for DMAPs and DMPs for AOs, ROSES, CANs.
- PDS does not evaluate DMPs.
- PDS does not develop pipelines for missions .
- PDS does not develop higher-order data products.

What PDS does not do (2)

- PDS does not actively participate in non-US mission data archiving without a NASA MOU or other agreement.
- PDS does not accept telemetry data.
- PDS does not archive executable software (see Finding XV).
- PDS does not curate geologic samples, analog materials, or physical objects.

LPSC Lunar and Planetary Science Conference 48th March 20–24, 2017 The Woodlands, Texas #LPSC2017

A New Planetary Data System Roadmap Study for 2017 – 2026

R. L. McNutt Jr.[1] , T. H. Morgan[2]

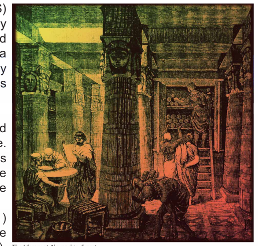
1 Johns Hopkins Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, MD,ralph.mcnutt@jhuapl.edu

2 Goddard Space Flight Center, Greenbelt, MD, thomas.h.morgan@nasa.gov .

Background: NASA established the Planetary Data System (PDS) in 1989 to deal with concerns that the data being returned by scientific satellites was in danger of being lost. The PDS, sponsored by NASA's Science Mission Directorate, archives electronic data products from NASA planetary missions. The PDS actively manages its archive to maximize usefulness, and the PDS has become a basic resource for scientists around the world.

All PDS-curated products are peer-reviewed, well documented, and available online to scientists and to the public without charge. Online search capabilities are also provided. The PDS uses ISO-based standards for describing and storing data that are designed to enable future scientists, who are unfamiliar with the original experiments, to analyze the data.

The PDS is organized as a federated data system (Figure 1) following the recommendations of the National Academy of Science Committee on Data Management And Computation (CODMAC). Scientist-led Discipline Nodes are organized around 6 broad areas:



The Library at Alexandria, Egypt

- **Atmospheres** (composition, structure, meteorology, and aeronomy) of planets;
- **Geosciences** (geology, geophysics, surface properties, and tectonics) of planets;
- **Small Bodies** (comets, asteroids, dwarf planets, and dust);
- **Planetary Plasma Interactions** (solar wind-planetary interactions, magnetospheres, ionospheres, plasma tori) of planets;
- **Ring-Moon Systems**; and
- **Cartography and Imaging Sciences** (pushbroom imagers, hyperspectral imagers, analysis tools) of solar system objects.

In addition, the PDS has two technical Support Nodes:

- The **Engineering Node** (systems engineering support, standards, technology investigations, coordination and development of system-wide software, and operations),
- NASA's **Navigation and Ancillary Information Facility** (SPICE, the observation geometry information system widely used by NASA planetary missions).

There is a small Project Office at Goddard, which manages funding and budgets, and establishes common policies across the PDS.

The PDS operates as a "living archive" of more than 1 petabyte of planetary data managed by subject-matter experts at the Discipline Nodes and used by scientists around the world. The PDS archive is constantly expanding to allow new mission data to be made available to the planetary science community. With the complex data provided by increasingly sophisticated spacecraft and instrumentation, there is a need to review the goals and objectives of the PDS.

PDS Roadmap Study: NASA's Planetary Science Division established this study to look ahead to the 2017-2026 time period. The study began in October of 2015 with the release of a Request for Information (RFI) asking for community input. The goal is to "develop a practical, community-developed pathway to implement the new long-term vision for the PDS, which continues to accomplish NASA's broad objective for the PDS; namely, preserving and making available all data products from planetary exploration research and missions".

A Roadmap Study Team (RST) consisting of individuals of differing backgrounds and interactions with the PDS was tasked to consider what steps the PDS should take during the next decade to progress. Initial examination began with the items in the RFI, namely, an examination of:

1. Tools, resources, workflows, tutorials, and interfaces
2. Making the archiving process seamless, less costly, and more efficient
3. The role of PDS relative to other archiving alternatives (e.g., journals), in providing the public access to NASA-generated data
4. Integration of PDS data products and services with those of other facilities, e.g., the U.S. Geological Survey's cartography program and the Minor Planets Center, and other products
5. The role the PDS should play in encouraging the development of higher-order data products
6. Appropriate improvements to the current search capabilities of the PDS

The Roadmap activity has included 19 telecons and 2 Face-to-Face meetings to date.

The Roadmap Team includes: Amitahba Ghosh (Tharsis Inc.), Anne Raugh (UMD), Denton Ebel (AMNH), Emily Law (JPL), Ernest Bowman-Cisneros (ASU), Flora Paganelli (APUS), Katherine Crombie (Indigo Information Services, LLC), Lisa Gaddis (USGS/Flagstaff), Matthew Tiscareno (SETI), Paul Ramirez (JPL), Renee Weber (MSFC), Ross Beyer (ARC/SETI), Thomas Stein (Wash U.), Thomas Morgan (GSFC), Ralph McNutt (APL), and Maria Banks (GSFC).

Results are centered on findings. These findings address: Progress relative to the goals of the last Roadmap; Changes in user expectations; Data discoverability and usability; Tools and file formats (including tools to translate PDS archive files to analysis-friendly formats); The implications of increased data volume, data complexity, and user numbers; Archiving laboratory data and data from samples; Including astromaterials data); Documentation and training; Data set citation, and information technology.

A draft Roadmap Report is expected by the end of April; The final by the end of May 2017.

Establishment of the Roadmap Activity

- NASA chartered this current PDS Roadmap Study Team to “develop a practical, community-developed pathway to implement the new long-term vision for the PDS, which continues to accomplish NASA’s broad objective for the PDS; namely, preserving and making available all data products from planetary exploration research and missions.”

Roadmap Study and Report

Planetary Data System (PDS)

Roadmap Study for 2017 – 2026

Contents

Activities Guided by Findings

- The Team formulated issues for the PDS for the next decade in the form of 16 “Findings”
- Some of these have identified “Remediations” and some note good progress with the need for the same
- The “Findings” are considered by the RST as objective facts that were developed over the course of extensive deliberations

Status of Findings

- There is consensus amongst the Team members on 13 of the findings, including the facts surrounding their importance, their statement, and possible actions that could be contemplated on responding to them
- There may be one more – likely on governance of the PDS
- The goal is to reach consensus on all of these

Finding 0

- While all PDS stakeholders are recognized as valuable, the prioritization of stakeholder interests and the impact those interests should have on PDS policy, design, and resource allocation are not clear.

Finding I

- There is a mismatch between the services and functions PDS is equipped to provide and the very high expectations of its users and NASA Management.

Finding II

- The PDS does an excellent job of providing access to its data holdings and is on track to increase such access. The latter is enabled by the PDS4 uniform metadata standard.

Finding III

- The accessibility and discoverability through the PDS4 metadata registry is a cornerstone to the future of community interaction with the PDS as a coherent storehouse of data. Legacy data archived in PDS3 format (the vast majority of PDS holdings) often lack metadata sufficient to enable discovery and accessibility commensurate with PDS4.

Finding IV

- There is a need of more translation programs that transform data from the PDS4 archive file formats to more usable analysis-ready formats.

Finding V

- It is a matter of concern whether the PDS nodes will have the resources to serve the data archiving requirements of individual ROSES investigations.

Finding VI

- A large amount of data from laboratory analyses of samples obtained by NASA missions is not archived and is in danger of loss. The status of astromaterials data archiving today is analogous to where planetary data archiving was before the PDS was chartered.

Finding VII

- A large amount of data from laboratory analyses of meteorites and cosmic dust is not archived and is in danger of loss.

Finding VIII

- The PDS4 information model is well-documented at a highly technical level. However, there is a critical need for broader documentation and training for all levels of users.

Finding IX

- There is a need for PDS to both expand and focus its search services, with a view to making it easier for users to find and execute the search appropriate to their query.

Finding X

- The PDS serves as the model for other national space mission data archives in ensuring future universal accessibility and searchability. The PDS is uniquely poised to lead efforts to make national and global archives interoperable.

Finding XI

- PDS is actively involved in addressing the data citation issue, and is well-positioned to provide the essential links in the chain needed to enable clear, direct referencing of PDS products; but it cannot itself change the habits and attitudes of authors, referees, and journal editors when it comes to including data set references in publications.

Finding XII

- **Higher-order products produced by mission teams beyond what is in their original data management plans are extremely valuable additions to the archive, but are not always included due to lack of resources needed by the mission to complete the archiving process.**

Finding XIII

- **The PDS has been and continues to be proactive in investigating information technology and adopting best practices.**

Finding XIV

- This Roadmap Study was initiated in the year immediately following a recompetition of the PDS Nodes, and will be completed at least 3 years and perhaps longer before the next recompetition, which limits the impact of a Roadmap Study activity on shaping the work of the PDS.

Finding XV

- The PDS is not an appropriate archive or repository for software.

Appendices

- The Appendices contain supporting material considered by the RST over the course of its deliberations
- Much of this material is not easily accessible otherwise
- These are included both to support the conclusions and statements of fact reached by the RST in support of its findings and to promote the transparency of that effort

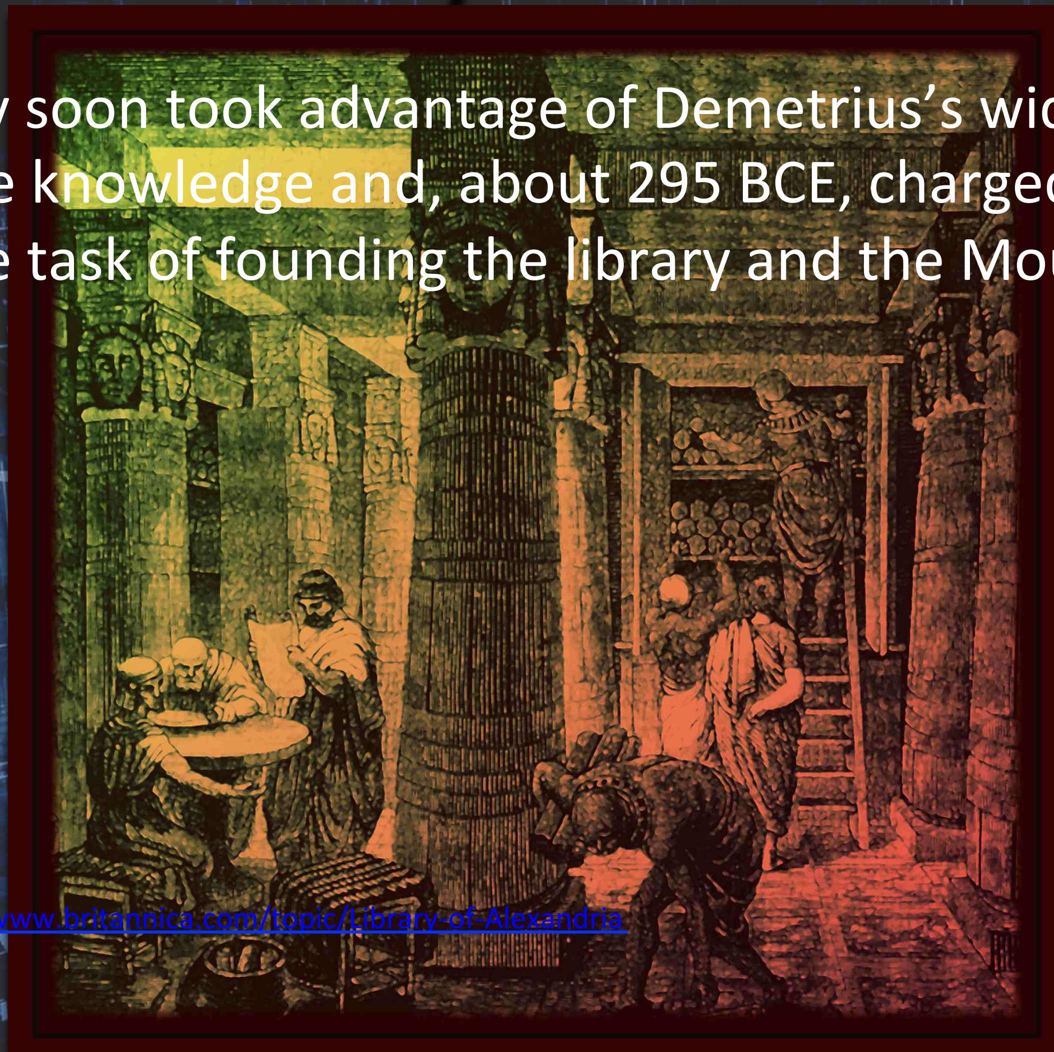
Challenge for the Next Decade

- As with any such large endeavor, there have been both known and unknown challenges, exacerbated with sometimes marginal budgets and the overall – and continuing – rapid evolution of IT infrastructure and how it impacts not only the U.S. digital world but the increasingly interconnected, and world-wide, “digital universe.”
- Such changes, in hardware, software, and middleware continue to evolve.
- Whether new approaches such as hyper-convergence or others yet to be defined will set the new infrastructure will be driven by the IT innovators, dealing with data access, storage, and manipulation on much large scales than of relevance here.
- Nonetheless, it is in the best interests of the PDS, and indeed of all of NASA’s SMD Data Centers to remain early adopters, so as not to be left behind. This requires continued vigilance and investment, both in people and technology.

The Final Word –
From the Conclusion / Summary

The Planetary Data System continues to fulfill a vital role in the preservation and curation of data obtained at great expense and with great effort from bodies other than Earth within the solar system. The risk of these data being lost forever, faced some 35 years ago has been averted. Upgrading PDS to PDS4 over the past decade has further enhanced the possibility of future preservation and use for the long term by securing the data collection to well-known and valued international standards for knowledge preservation and description.

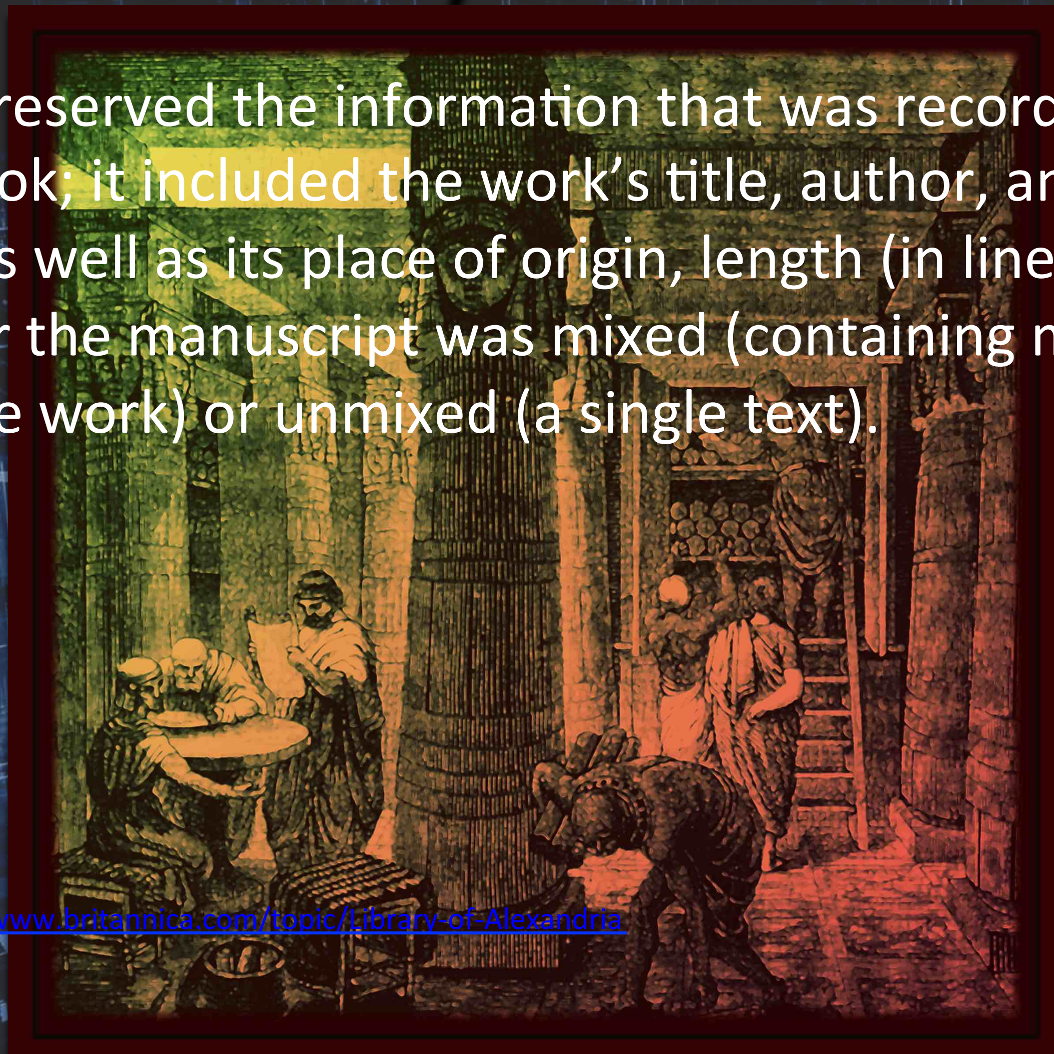
The Library of Alexandria



Ptolemy soon took advantage of Demetrius's wide and versatile knowledge and, about 295 BCE, charged him with the task of founding the library and the Mouseion.

From https://www.britannica.com/topic/Library-of-Alexandria

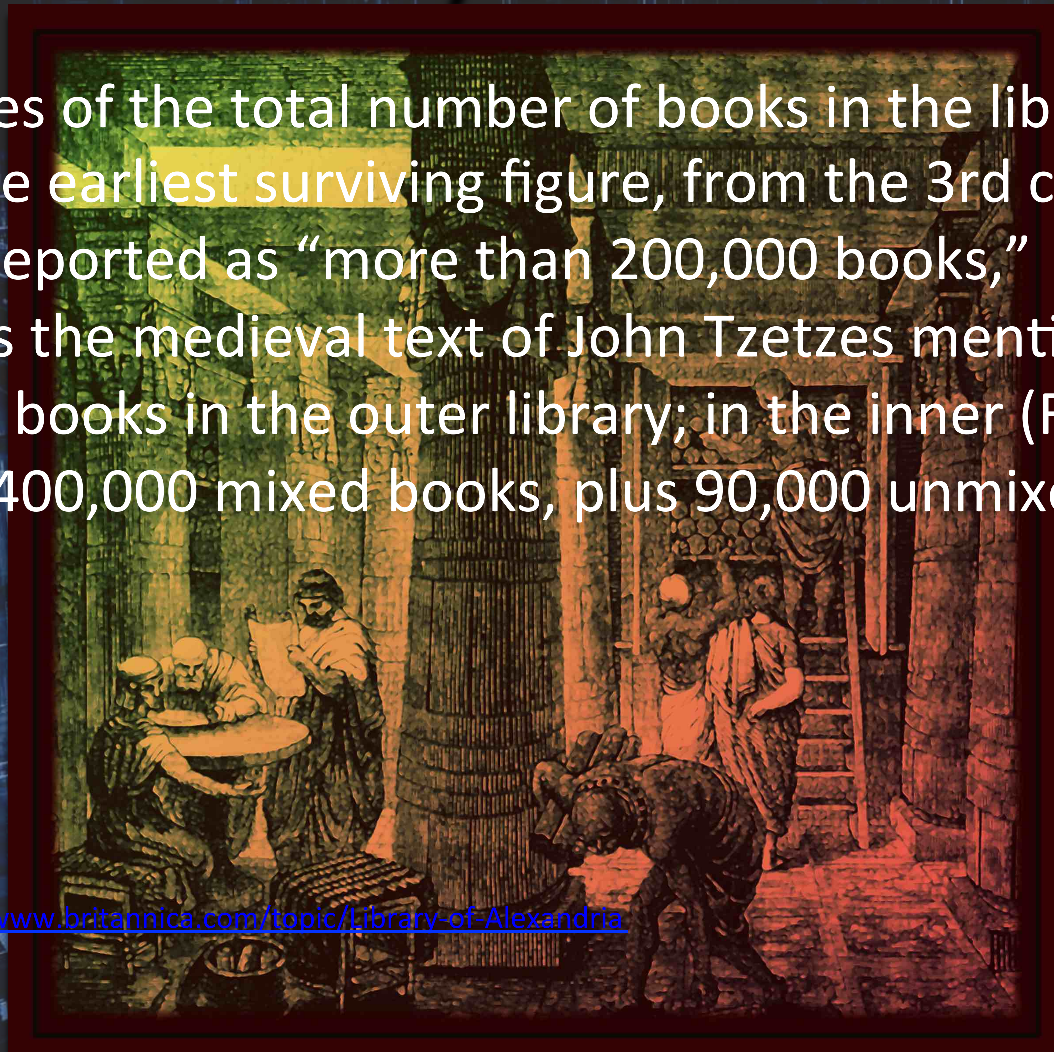
The Library of Alexandria



Galen preserved the information that was recorded for each book; it included the work's title, author, and editor as well as its place of origin, length (in lines), and whether the manuscript was mixed (containing more than one work) or unmixed (a single text).

From https://www.britannica.com/topic/Library-of-Alexandria

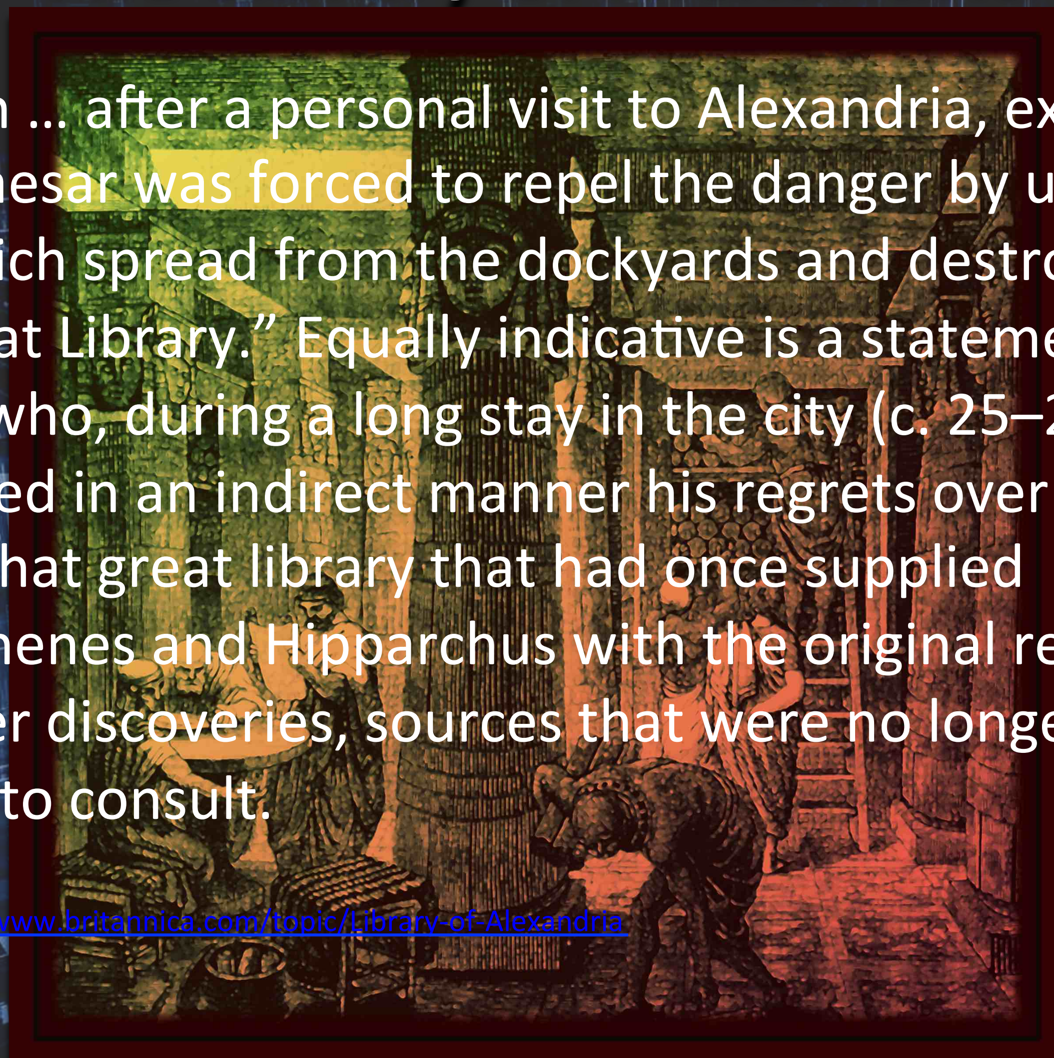
The Library of Alexandria



Estimates of the total number of books in the library vary. The earliest surviving figure, from the 3rd century BCE, is reported as “more than 200,000 books,” whereas the medieval text of John Tzetzes mentions “42,000 books in the outer library; in the inner (Royal) Library 400,000 mixed books, plus 90,000 unmixed books.”

From https://www.britannica.com/topic/Library-of-Alexandria

The Library of Alexandria



Plutarch … after a personal visit to Alexandria, explained that "Caesar was forced to repel the danger by using fire, which spread from the dockyards and destroyed the Great Library." Equally indicative is a statement by Strabo who, during a long stay in the city (c. 25–20 BCE), expressed in an indirect manner his regrets over the loss of that great library that had once supplied Eratosthenes and Hipparchus with the original reports of earlier discoveries, sources that were no longer there for him to consult.

From https://www.britannica.com/topic/Library-of-Alexandria