



STEWARDED THE INVISIBLE

Setting the stage for Institution-wide digital preservation at the Smithsonian

Digital Preservation Assessment Final Report
Executive Summary Presentation



November 15, 2016

Kara Van Malssen (kara@avpreserve.com)

Chris Lacinak (chris@avpreserve.com)

ASSESSMENT GOALS

1. Articulate risks to digital collections and research data across the Smithsonian Institution
2. Identify gaps in stewardship of digital assets
3. Situate digital preservation in strategic vision & initiatives
4. Establish common vocabulary & reference points
5. Provide recommendations to address gaps, risks, and reach SI-wide goals



Interviews

With 16 stakeholder groups representing collecting units, data repositories, and content creators



Survey

On data totals and storage practices of researchers across the SI — 100 complete responses were submitted



Documentation Review

Of pertinent strategy, policy, analysis, and other relevant SI documentation

METHODOLOGY

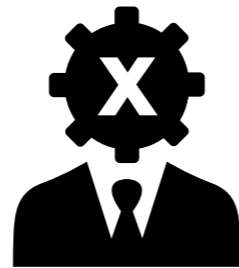
SUMMARY OF FINDINGS

There are significant digital stewardship gaps today that place assets at risk and use available resources inefficiently



Existing policies put preservation responsibility exclusively on resource creators

No central supporting roles are articulated by SD 610. There is no official policy for research data management. There is no central digital preservation oversight, infrastructure, guidelines or other resources.



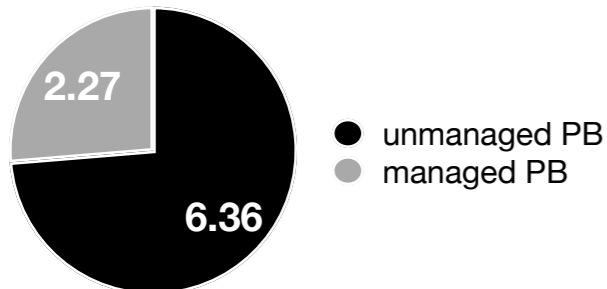
Content creators do not have the resources, expertise, and at times, incentives, to ensure that digital assets persist

Researchers and digitization project managers default to paths of least resistance—store research data on hard drives and put collections and institutional output in the DAMS. Metadata, essential to longevity, is often limited or not created.



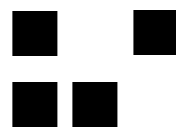
The current project-centric approach to data management is inefficient and risk prone

Research data management plans and SI DAMPs require project-centric solutions. Without oversight to ensure alignment with SI interests, economies of scale and scope are lost, and resources often neglected after the project ends.



There is now ~2 PB of collections data, and ~6.7 PB of research data across the SI

While >75% of collections data is in a managed environment like the DAMS, >85% of research data is unmanaged and ~35% is not under SI control. Rapid digitization and data-driven research will increase these numbers >55 times by 2026.



Current repository infrastructure is lacking in capacity and functionality

The available infrastructure cannot support the volume of data across the SI that staff have identified as valuable. Repositories available today for research data can only accommodate ~1% of data that researchers want to keep indefinitely.



Lack of a digital preservation mandate, undefined responsibilities, and unclear definitions exacerbate inaction

Excellent work is being done by many members of the Digital Preservation Working Group, but these are siloed efforts. Organizational alignment is required to maximize benefit at the enterprise level.

RECOMMENDATIONS

Make digital preservation an underlying, systematic function of the Smithsonian Institution

1

Instill a sense of urgency amongst stakeholders

Quantify the need and communicate it broadly

IMPACT ☆ Identify champions at all levels, galvanize broad participation

2

Establish governance & oversight

Create a Digital Preservation Directorate and a Digital Preservation Advisory Board, and define cross-institutional roles and responsibilities down to the level of the individual researcher

IMPACT ☆ Defined responsibilities empower stakeholders to act

3

Create a vision for digital preservation

Demonstrate the value and impact of digital preservation to achieving SI's Vision. Incorporate the role of digital preservation into the next strategic plan

IMPACT ☆ Digital preservation becomes a shared objective of the organization

4

Create & update policies

Update and formalize terminology. Create an SD for digital preservation, and another for research data management

IMPACT ☆ Resolve questions and establish reference points

5

Ensure alignment & accountability

Conduct training and outreach, align formal terms to discipline-specific verbiage, track metrics

IMPACT ☆ Consistent communication establishes feedback loops & engenders trust

6

Ensure availability of technical infrastructure

Technical resources should reflect mandates and support functional needs of researchers and collecting units

IMPACT ☆ Mandates and policies become achievable

7

Ensure sustainability

Formalize programmatic funding streams that are supplemented by project and grant funds

IMPACT ☆ Eliminating reliance on short-term funding enables long-term management

8

Implement a phased approach

Systematize data management practices moving forward, tackle the backlog later

IMPACT ☆ Forward progress is enabled by a focus on actionable steps and demonstrable wins