PROCURING IT SYSTEMS

Thinking about digital preservation from the start

Michael Popham

Digital Preservation Coalition
UK
michael.popham@dpconline.org
0000-0002-6842-4294

Jenny Mitcham

Digital Preservation Coalition
UK
jenny.mitcham@dpconline.org
0000-0003-2884-542X

Paul Wheatley

Digital Preservation Coalition
UK
paul.wheatley@dpconline.org
0000-0002-3839-3298

This poster presents one of the final outputs produced from a collaborative project between the Digital Preservation Coalition and the Nuclear Decommissioning Authority in the UK, "Digital preservation requirements for procuring IT systems". This addition to the Digital Preservation Coalition's Procurement Toolkit proposes the requirements that should be considered when procuring an IT system (for example an EDRMS, DAMS, or GIS) that may ultimately contain at least some records or digital content that needs to be retained beyond the life of the system.

Keywords – Procurement, IT systems, data export
Conference Topics – Sustainability: real and imagined; From Theory to Practice.

I. INTRODUCTION

This poster will summarize the set of six key requirements to consider when procuring any IT system which might contain data of long-term value. It was developed to inform IT procurement practices at the UK's Nuclear Decommissioning Authority (NDA) and was published as an addition to the Digital Preservation Coalition's (DPC) existing Procurement Toolkit [1].

By enabling content extraction in a managed way from IT Systems, organizations can avoid costly barriers to preservation and/or migration of content when an IT system is retired at the end of its life and ensure continued access to valuable information.

The DPC's Procurement Toolkit is intended to provide straightforward advice on how to get the best result out of a procurement process. The bulk of the Toolkit focuses on the procurement of a digital preservation system, but the joint NDA-DPC project "Reliable, Robust and Resilient Digital Infrastructure for Nuclear Decommissioning" (2019-2023) [2] both highlighted the challenges of accessing and secure critical data in legacy systems and aimed to enable the NDA to commission future data and systems with long term resilience from the outset. It seemed helpful to develop some general principles which could be applied to the procurement of *any* IT system which might contain data of lasting value, and it was apparent that advice such as this would benefit many other organizations as well as the NDA.

II. USAGE

The result was a set of six proposed statements of requirement [3] which might usefully be incorporated into the procurement of any IT system – rather than one which was specifically intended for digital preservation. That being said, we encourage users to adapt the language used in these statements to match their organization's particular circumstances, and also to add additional requirements as necessary. These requirements are

iPRES 2023

underpinned by some basic principles, and also accompanied by a statement emphasizing the importance of acceptance testing; these are repeated below.

III. PRINCIPLES

The following principles should be applied to ensure that any content ultimately selected for long-term preservation is managed effectively before transfer to a digital preservation system:

- Appropriate records management policy and procedure should be put in place and fully documented, including clear criteria and related processes for record disposal, retention, and long-term preservation.
- Data and metadata should be structured in a way that makes it straightforward to use and re-use beyond the life of any particular IT system. Open data standards, metadata standards and file formats that facilitate data interoperability are encouraged.
- Robust processes for backing up current data should be applied.

IV. ACCEPTANCE TESTING

It is essential to verify that a product selected during a procurement process does in fact meet the specified requirements in practice. The inclusion of a content import, extraction, and preservation scenario as part of user acceptance testing is therefore recommended. Testing with a sample set of content and metadata that has been extracted from the IT system can be useful in flagging up any issues before it is too late to make significant changes.

V. REQUIREMENTS

The six statements of requirement are given below, each with an associated rationale:

1. **The system should use appropriate open data standards to structure and store data.**
Rationale: Data standards facilitate subsequent data interchange and interoperability without the need for costly and/or complex data migration.
2. **It must be possible to import and store content and associated metadata, if the system is to be populated with existing data.**
Rationale: If the system to be procured will initially be populated with data from an existing system that it is replacing then it will be necessary to ensure that the data as well as accompanying metadata can be effectively imported and stored.
3. **The system should enable digital content to be selected for disposal or retention/preservation as appropriate:**
a. Flagging of content by users for action or for specific retention periods.
b. Selecting content for extraction using search on content and/or metadata.
Rationale: Not all content held within a system will be of equal value or will need to be kept for the same period of time. Being able to manage retention periods and mark content for deletion or for preservation are important features to help ensure that the right content is managed for the right period of time.
4. **The system must provide a practical mechanism for the extraction of digital content, such as via an API and/or user interface.**
Rationale: An IT system has a finite lifespan. Suitable export options must be available if the content held within the system has a retention need beyond the life of the system itself.
5. **The system must enable appropriate metadata, structural and contextual information to be extracted along with the digital content.**
Rationale: Digital content may be of little value without metadata that helps it to be located, understood, and trusted.
6. **The system must allow the extraction of digital content and metadata in formats that will permit its use outside of the system.**
Rationale: Dependence on an obsolete IT system may hamper or prevent the understanding and use of digital content and metadata.

VI. CONCLUSION

Presenting these requirements as a poster at iPres 2023 will open them up to scrutiny and adoption by a wider audience, and will empower digital preservation practitioners to provide valuable input into procurement processes within their own organizations to ensure that digital preservation requirements are factored into IT system procurement.

1. REFERENCES

[1] https://www.dpconline.org/digipres/implement-digipres/procurement-toolkit

[2] https://www.dpconline.org/digipres/collaborative-projects/nda-project

[3] https://www.dpconline.org/digipres/implement-digipres/procurement-toolkit/procurement-toolkit-it-systems