



THE NATIONAL ARCHIVES
DIGITAL PRESERVATION DEPARTMENT
PRONOM 3
USER REQUIREMENTS

Release: 0c

Date: 19 June 2003

Author: Adrian Brown, Digital Archives Analyst

Owner: Digital Preservation Department

Client: The National Archives

Document Number: PRONOM 3-URP-01

Document Control

Author: Adrian Brown, Digital Archives Analyst

Approved By:

"I have read this document and agree that it addresses adequately the business needs and technical requirements of the National Archives".

David Ryan, Head of Archive Services.

Document Reference: PRONOM 3-URP-01

Issue: 0c

Issue Date: 19 June 2003

Source File: I:\Digital Preservation\IT Projects\PRONOM File Format system\File format history - Pronom 3\PRONOM3 User Requirements 1a Public.doc

Distribution

Public

Document History

Issue	Date	Author	Comments
0a	11 June 2003	A. Brown	First draft
0b	14 June 2003	A. Brown	Second draft
0c	19 June 2003	A. Brown	Public consultation draft

Contents

1	PURPOSE OF DOCUMENT	4
1.1	BACKGROUND.....	4
1.2	BUSINESS PROCESSES.....	4
2	PROJECT OBJECTIVES	5
3	REQUIREMENTS	6
3.1	NON-FUNCTIONAL REQUIREMENTS.....	6
3.1.1	<i>Supplier</i>	6
3.1.2	<i>Interface</i>	6
3.1.3	<i>Security</i>	6
3.1.4	<i>Performance</i>	6
3.2	FUNCTIONAL REQUIREMENTS.....	6
3.2.1	<i>Environment</i>	6
3.2.2	<i>Data Structure</i>	6
4	REQUIRED ENHANCEMENTS	7
4.1	WEB-ACCESS.....	7
4.2	PROVENANCE INFORMATION.....	7
4.3	VENDOR INFORMATION.....	7
4.4	PRODUCT IMAGES.....	7
4.5	INTERFACE.....	7
4.6	SEARCH FUNCTIONALITY.....	8
4.7	SEARCHES.....	8
4.8	RESULT SETS.....	9
4.9	REPORTS.....	9
4.10	DOCUMENTATION.....	9

1 Purpose of Document

The purpose of this document is to define the user, reporting and technical environment requirements of PRONOM 3. PRONOM is a database designed to store information on software applications and their related file formats.

A full review and understanding of these requirements will result in a Functional Specification document.

1.1 Background

PRONOM is an application for managing information about the file formats used to store records (including the hardware and software environments in which records were created) in government departments. This will be used to support the long-term preservation of the records themselves after they have been transferred to the National Archives (TNA) for permanent storage.

PRONOM was produced by Tessella Support Service Ltd in March 2000. The current version is Version 2, received in November 2002 for user acceptance testing. This incorporates support for foreign languages.

1.2 Business Processes

PRONOM is a software tool which is required to support the following TNA business processes:

- Accession: PRONOM will be used as part of the accession process for electronic records into the digital archive. It serves three main functions: firstly, to automate the identification of the record file formats, secondly, to automate metadata extraction at the file level, and thirdly, to act as a central repository for information about file formats and their technical dependencies, thereby eliminating the need to duplicate this information within the digital archive. PRONOM will be fully integrated into the accession workflow, identifying formats or reporting exceptions requiring manual identification, extracting metadata, and exporting that information into the digital archive.
- Preservation: PRONOM will support long-term preservation, informing specific strategy decisions through the automated generation of migration pathways. It will be integrated into the Technology Watch process.
- Delivery: PRONOM will support the delivery of electronic records to users, by enabling the automatic configuration of the appropriate software environments. This feature will be developed further in a future version.
- In addition, the release of PRONOM on the Internet will provide an important resource to the broader digital preservation and IT communities, by enabling the dissemination of authoritative information on file formats and software products.

2 Project Objectives

The overall project objectives are to develop a fully operational product that:

- Contains a repository of information about file formats, software products, database environments and operating systems used in government departments to generate records.
- Supports the automatic identification of file formats and extraction of metadata through examination of the file structure.
- Reports the life cycles of software products and file formats for analysis.
- Supports multi-lingual versions through foreign language field prompt replacement.
- Is accessible on the World Wide Web.
- Can be updated either for program functionality or data or both from an internet download or through a new CD ROM sent to a user.
- Has the reporting capability to trigger and inform reviews of records in file formats that are no longer supported by current products or records produced using products that are no longer supported by their supplier.
- Supports the import of information from other sources, such as the NIST National Software Reference Library Reference Dataset (NSRL RDS).

The objective of Phase 3 of the project is to create a web-enabled version of PRONOM.

3 Requirements

3.1 Non-functional requirements

3.1.1 Supplier

1.	The supplier must visit National Archive premises at Kew to install and configure the software
----	--

3.1.2 Interface

2.	System User Interface must be web based GUI.
3.	System User Interface must be intuitive and easy to use by all trained staff.

3.1.3 Security

4.	Access to maintenance screens must be authenticated through password control (e.g. MS Windows 2000 logon).
----	--

3.1.4 Performance

5.	The system must be able to handle one hundred concurrent users with no detrimental effects on performance.
----	--

3.2 Functional requirements

3.2.1 Environment

6.	The National Archives operates Windows 2000 operating systems for desktop and server.
7.	The read only version of this system will be accessible by all users at Kew and on the WWW and therefore must be able to be run in a web browser.
8.	The read only version must be cross browser compatible and operate in any browser that operates in accordance with W3C.org .
9.	The system must be year 2000 compliant.
10.	The system must comply with e-GIF version 5.

3.2.2 Data Structure

12.	The data must be stored in a recognized database structure.
13.	The database must be capable of export to a commonly agreed format in accordance with e-gif, including XML and CSV.

4 Required Enhancements

This section describes only those changes which will be required to the functionality of the system. All functionality not mentioned here must be retained.

TNA will determine an appropriate sub-domain of the www.nationalarchives.gov.uk domain to host the PRONOM site. TNA will host PRONOM from existing TNA web-servers. TNA will provide a new homepage for the PRONOM site, incorporating links to the PRONOM system and subsidiary information pages.

4.1 Web-access

4.1.1	The system will be publicly accessible on the WWW, via a sub-domain of the TNA web domain (see above).
4.1.2	Installation of the system will be undertaken by the supplier in conjunction with the TNA web team.

4.2 Provenance information

4.2.1	Addition of database fields to record the provenance and currency of information recorded on PRONOM. These should comprise details of the information provider (e.g. name, organisation, address), and the date the information was added to PRONOM.
4.2.2	The provenance information will be added manually; the date should be inserted automatically by the system whenever a new record is created.
4.2.3	The user reports should be amended to display the provenance information.
4.2.4	The maintenance screens should be amended to allow users to add, edit, and delete provenance information.

4.3 Vendor information

4.3.1	Addition of database field to record the trading history of the vendor, e.g. changes of vendor name, mergers, buy-outs, with dates.
4.3.2	The vendor report should be amended to display this information.
4.3.3	The maintenance screens should be amended to allow users to add, edit, and delete this information.

4.4 Product images

4.4.1	The system will allow images of product packages to be linked to each product.
4.4.2	Supported image formats should include TIFF, JPEG, GIF and PNG.
4.4.3	Where an image is available, this should be displayed as part of the product report.
4.4.4	The system should allow multiple images per product

4.5 Interface

4.5.1	The look and feel of the interface must be easily changed by TNA. All pages will use a table-less layout, controlled by CSS. All pages will contain valid XHTML 1.1.
4.5.2	The look and feel of the interface will be modified to conform to the new TNA style. Style guidance will be provided by TNA

4.5.3	The content of the “homepage” will be modified. The new content will be provided by TNA.
4.5.4	The menu bar will be removed and replaced by a side bar entitled “Search by”, containing links to each search form.
4.5.5	The searches will be renamed as follows: Vendor, Product File Format, Release Date, Support Period.
4.5.6	The “Contact Us” section will be removed entirely – this information will be available from the PRONOM homepage.
4.5.7	The “Time to run” indicator on the search pages will be removed.
4.5.8	The timestamp which appears on all pages will be changed to long date format and repositioned (new position to be provided as part of style guide).

4.6 Search functionality

4.6.1	Search functionality will be simplified to provide a text box for users to enter a search string, and a “Search” button to execute the search. Date-based searches will provide pick-lists for day, month and year.
4.6.2	If no search term is entered, all records will be returned.

4.7 Searches

4.7.1	Vendor
4.7.1.1	Users will be able to perform a search by vendor name and return a list of all vendors with names containing the search string.
4.7.1.2	The returned result set will contain: Vendor Name. It will be sorted by Vendor Name. Each row will contain a link on Vendor Name to the relevant Vendor report.
4.7.2	Product
4.7.2.1	Users will be able to perform a search by product name and return a list of all products with names containing the search string.
4.7.2.2	Users will be able to perform a search by vendor name and return a list of all products with vendor names containing the search string.
4.7.2.3	The returned result set will contain: Product Name, Version, Vendor Name. It will be sorted by Product Name. Each row will contain a link on Product Name to the relevant Product report, and on Vendor Name to the relevant Vendor report.
4.7.3	File Format
4.7.3.1	Users will be able to perform a search by file format name and return a list of all products which support (read or write) file formats with names containing the search string. Users should have the option to search on all products, or to restrict the search to currently supported products.
4.7.3.2	Users will be able to perform a search by file format extension and return a list of all products which support (read or write) file formats with extensions matching the search string. Users should have the option to search on all products, or to restrict the search to currently supported products.
4.7.3.2	The returned result set will contain: Product Name, Version, Vendor, Read, Write, Invariance. It will be sorted by Product Name. Each row will contain a link on Product Name to the relevant Product report, and on Vendor Name to the relevant Vendor report.
4.7.3.4	Users will be able to perform a search by file format extension and return a

	list of file format names with extensions matching the search string.
4.7.3.5	The returned result set will contain: Format Name, Format Extension. It will be sorted by Format Name.
4.7.4	Support Period
4.7.4.1	Users will be able to perform a search by date and return a list of all products still supported on that date.
4.7.4.2	Users will be able to perform a search by date and return a list of all products not supported from that date.
4.7.4.3	The returned result set will contain: Product Name, Version, Vendor Name. It will be sorted by Product Name. Each row will contain a link on Product Name to the relevant Product report, and on Vendor Name to the relevant Vendor report.
4.7.5	Release Date
4.7.5.1	Users will be able to perform a search by date and return a list of all products with release dates prior to or subsequent to the search string.
4.7.5.2	Users will be able to perform a search by date and return a list of all products released between two dates.
4.7.5.3	The returned result set will contain: Product Name, Version, Vendor Name, Release Date . It will be sorted by Product Name. Each row will contain a link on Product Name to the relevant Product report, and on Vendor Name to the relevant Vendor report.

4.8 Result sets

4.8.1	The ability for users to resort result sets by different columns will be retained.
4.8.2	This functionality is not immediately apparent in the current interface – the existing hypertext link appearance of the column headings will be replaced by a visual indicator for the column on which the sort is currently active (e.g. a Windows-style downward arrowhead).

4.9 Reports

4.9.1	The existing report options (XML, CSV and printer-friendly) will be retained.
4.9.2	The style of the printer-friendly report will be modified in accordance with the new style guide.
4.9.3	The supplier will suggest an alternative on-screen display method for the product report, to minimise the need for users to scroll downwards. Options for separating the report into several pages or tabbed sections will be investigated.

4.10 Documentation

4.10.1	In addition to the standard project documentation, the supplier will produce a “plain-English” user guide for external users of the system. This will be provided in XHTML 1.1 and Word 2000 formats.
4.10.2	The supplier will provide copies of the updated XSD files for all XML schemas used in the system.