ArchivesSpace Reports Module

Requirements, Draft, 2017 May 9

The ArchivesSpace program aims to implement an interactive and customizable report module for users of the ArchivesSpace application. Both default and editable formatted reports and a generic tabular data export per each browse and search result screen are wanted.

Archivists must have the ability to define, modify, and save a report definition; save to a report library a new or modified report definition (to share); and add to a report library and use report definitions acquired from other users.

General specifications

- Must use a Free and Open Source Software solution.
- Must support report generation, including exporting tabular data from search and browse results and approximately 40 default report definitions.
- Must support CSV, PDF, RTF, HTML, or JSON report export formats.
- Must support modifying default reports and their reuse.
- Must support creation of new reports and their use.
- Must explicitly indicate the data fields selected for each report to give users understanding of the report's intent and results.
- Reports definitions should NOT include stored procedures, or where stored procedures are necessary the report definition should explicitly describe the procedure.
- All modified and new reports should be maintained across application updates.
- Reports should avoid redundancy, and instead rely on the ability to filter results by specified parameters for customization.

Report Permissions

- Generating reports and CSV exports requires permission be assigned to the user. Without such permission, a user will not be able to see the list of reports or the option(s) to export CSV data.
- Adding or deleting reports from the Reports Library requires permission be assigned to the user. Without such permission, a user will not have access to the reports library.

Generating Reports

- Generating a report needs to be configured to run as background jobs to avoid timeout occurrences.

- A generated report will be limited to the active repository; however database administrators should have the option to run a report across all repositories in a single ArchivesSpace implementation.
- There will be a Reports Selection List in the staff interface from which users will be able to select a report to generate.
- For each report, the user must be able to pass sort and filter parameters; include primary, secondary, and tertiary sorts and filters.
- A report will be able to be canceled after it has started and before it has completed.
- Users must be able to limit report queries by null values to support quality assurance and data clean up tasks.
- Reports generated should be able to be saved, edited, and rerun.
- Users should be able to pass sorts and filters to the data set before it is exported.
- Data exported as CSV must be in tabular format so as to allow archivists the ability to sort, filter, edit, and format the exported data and to enable the creation of summary charts and tables.
- Report data exported as CSV must display discrete values in separate columns. For example, `extent.portion_id`, `extent.number`, `extent.extent_type_id`, `extent.container_summary` should be reported in separate columns, not merged in a single field.

Download CSV Reports

- Users will be enabled to export CSV reports from any search or browse screen.
- The CSV export will include all fields in browse or search data set, that is, if the browse field is for Resources, then the CSV download will include all data fields for resource records and associated sub-records. However, the user should have the ability to omit fields from the data set.

Formatting Reports

- Users should be capable of generating reports as CSV, PDF, RTF, HTML, or JSON.
- All open source fonts and scripts should be supported. Proprietary fonts such as Arial should not be supported.
- Field labels should match the labels used in the staff interface of ArchivesSpace. For example, the label for `accession.provenance` should be Provenance.
- Report data exported as HTML must exclude HTML styles and classes, including colors, fonts, etc, thereby enabling easier use of data in a variety of content management systems.
- Enumeration identifiers should be automatically translated to the enumeration values. For example, if the user wants the extent type to display, the number 277 should be automatically translated to "linear feet" in the resulting export.

- Brackets, quotation marks, commas, and null values should be removed from reported identifier fields, including accession, resource, and digital object identifiers. Instead, subfields should be separated by hyphens, e.g. *ABC123-DEF456-GHI789-JKL123*.
- Report data exported in PDF and RTF formats must include footers that feature the title of the report, the date the report was generated, and the page number.
- Report data exported in PDF, RTF, and HTML formats must use appropriate heading tags (h1, h2, etc.) to facilitate accessibility.

Reports Library

- All pre-defined reports will be accessible in a reports library (folder) that is part of the ArchivesSpace folder stack.
- Reports subsequently modified or created can be added to the Reports Library by user(s) having the required permissions.
- All reports in the Reports Library should be capable of being added to or removed from the Reports Selection List. A process will exist for adding / removing reports from the Report Selection List.
- The Reports Library will be accessible only to users having assigned permissions to add or remove reports from the Reports Library.

Modifying Report Definitions

For any report in the Reports Library and registered in the Report Selection List, a user must be able to
- remove fields from the report definition;
- add fields to the report definition from the tables already in use by the report definition;
- and change the arrangement and formatting imposed on the results.