

ArchiveDb

Part 1

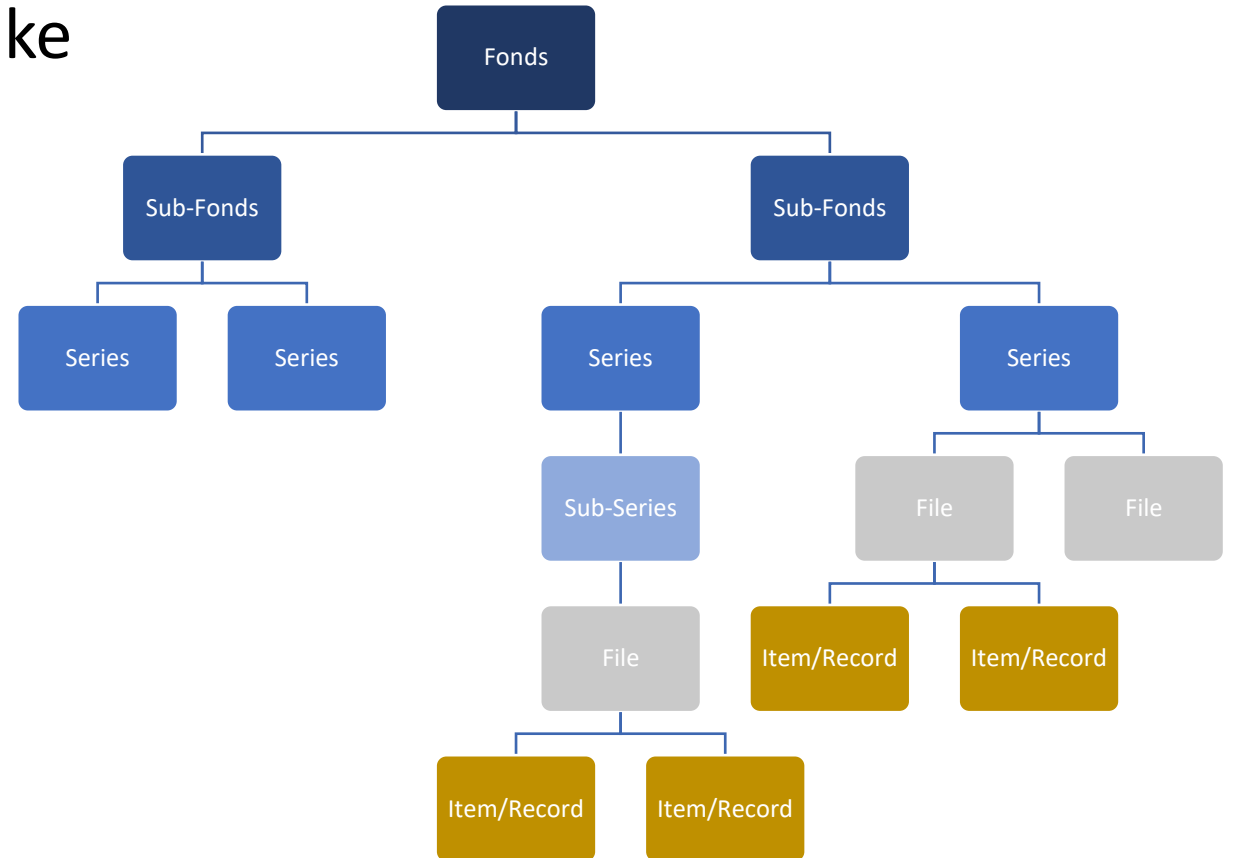
Introduction

- Falmouth History Archive had been digitising its archive to spreadsheets and Microsoft Access databases for some years prior to about 2017 but needed a way to integrate these for unified searching
- ArchiveDb began as a solution to this requirement of a low-cost, multiuser, database capable of importing these diverse existing data and providing that comprehensive, unified, search, including full text search of attached content
- Ongoing digitization continues to create diverse tabular & list data, PDFs, images, etc., requiring flexible import function for future data

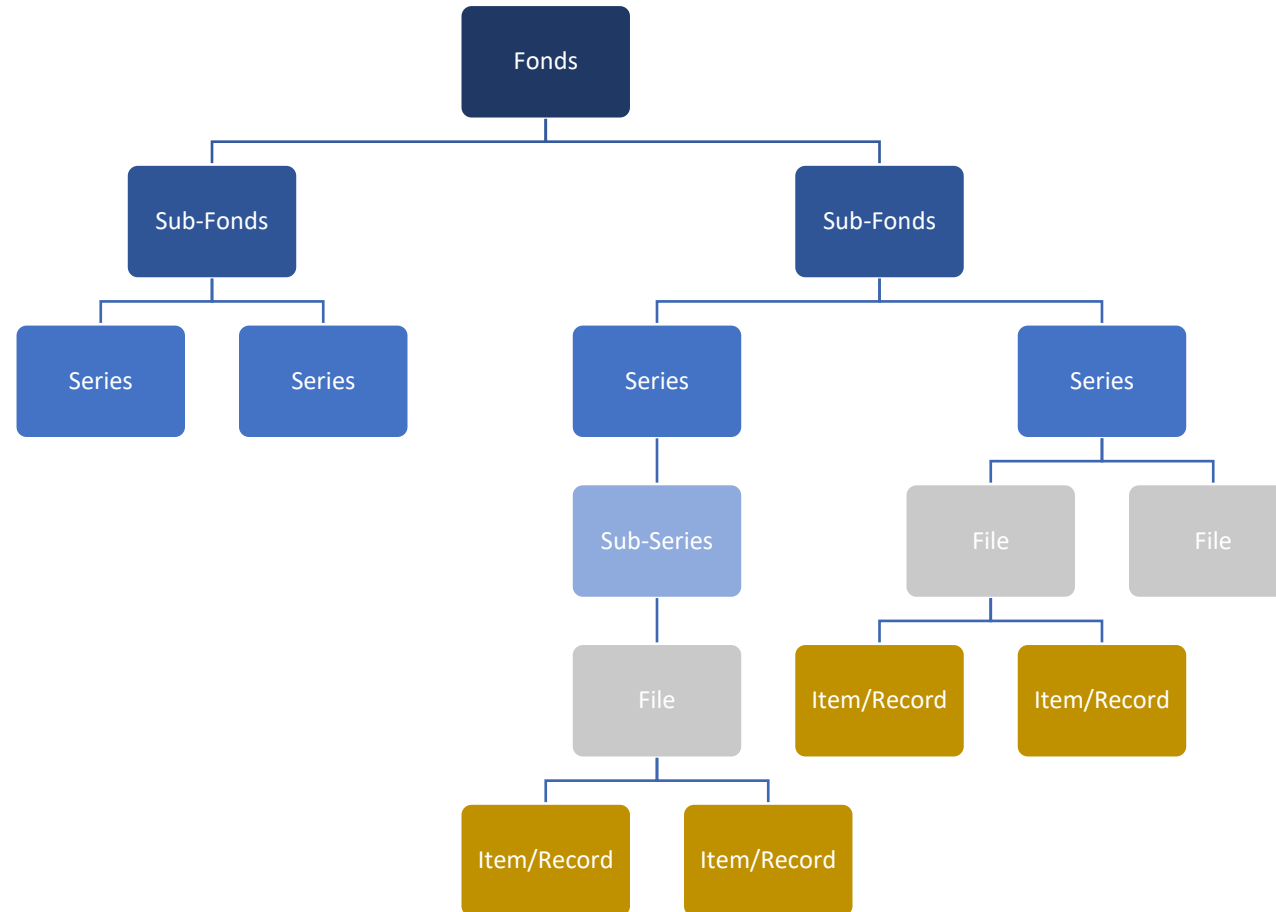
- ArchiveDb serves as both an information resource and a collection management system
- It offers:
 - A one stop full text search of the entire database, including attached content
 - Comprehensive/detailed searches of all other data types
- ArchiveDb follows standard archival hierarchical principles conforming to ISAD(G)
- To cover broad use cases, ArchiveDb is designed with flexible configuration

Archival Hierarchy

- Data is organised in a tree structure like the directories on a computer
- This allows compliance with *General International Standard Archival Description*:
 - Fonds
 - Sub-fonds
 - Series
 - Sub-series
 - File
 - Item



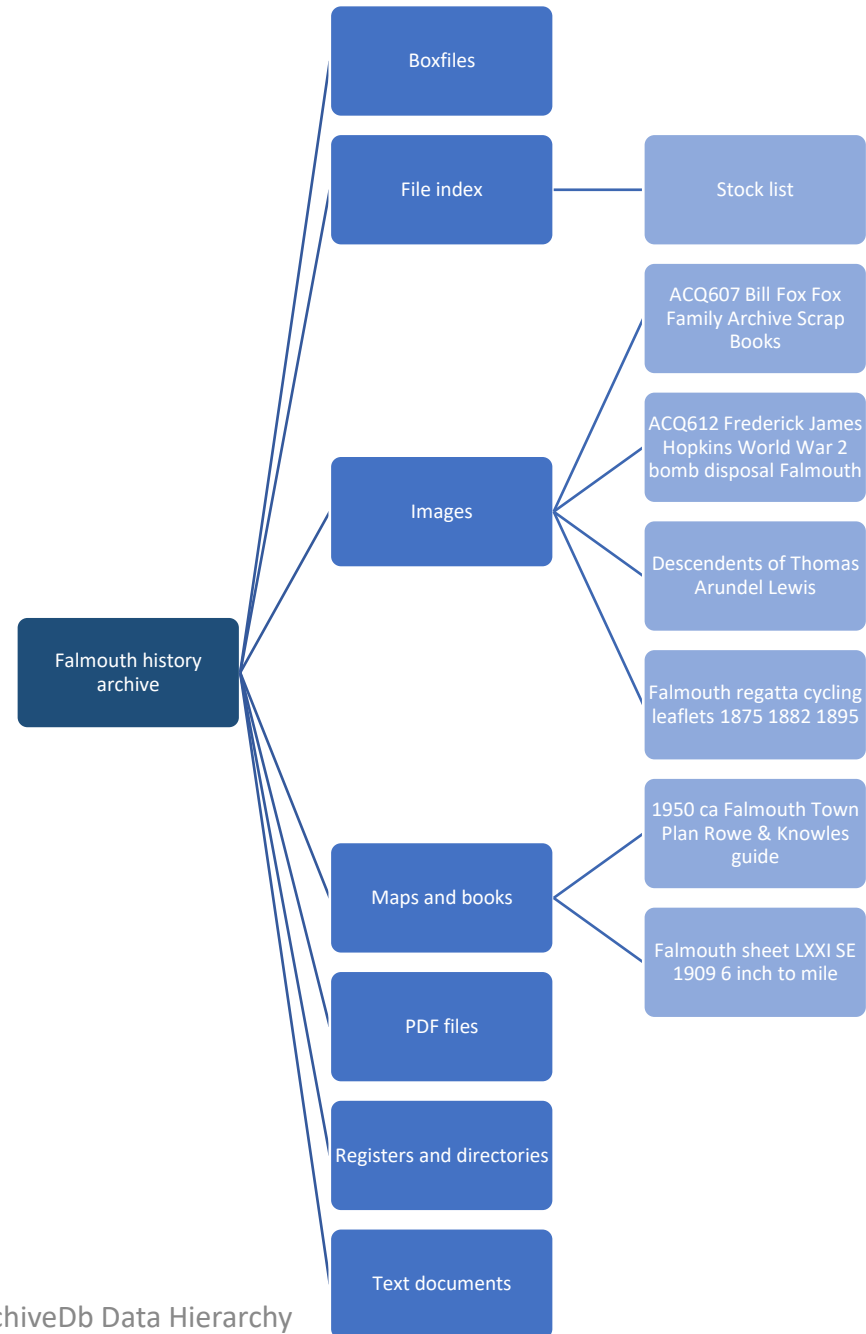
Archival Hierarchy



Next: Layout of Falmouth Archive


Layout of Falmouth History Archive hierarchy (incomplete)

The hierarchy is displayed vertically in the ArchiveDb user interface




Next: ArchiveDb Data Hierarchy

ArchiveDb Data Hierarchy

- The root node in ArchiveDb represents the **Institution**
- All other nodes are children of the root node and represent **Fonds, Sub Fonds, Series, Sub-Series or Files**
- In ArchiveDb these are colour coded according to level of description
- ArchiveDb **Files** are children of sub fonds, series or sub-series
- Files icons have 3 horizontal blue bars 
- Files are containers for **Records**, (or ISAD(G) Items)
- Files display records in tables

Files, Records and Fields

- Each file contains a table:

BF11 Penryn 						
Serial	BF Number	BF Serial No	Title/Description	Date/Period Covered	Document Type	Remarks
1	Penryn 1 Box 11	11.1	The story of the foundation	1845 - 1848	Typed A4 Document	2 copies
2	Penryn 1 Box 11	11.2	Jubilee of 150 years of Not	1845-1995	Souvenir Booklet	2 copies
3	Penryn 1 Box 11	11.3	Penryn versus Hilton Cleav	1953-1954	Souvenir Programme	
4	Penryn 1 Box 11	11.4	Seventy years of St.Glurias.	1908-1982	Document	
5	Penryn 1 Box 11	11.5	Tremough School. Small as	1976-1982	Brochures and Magazine	Blue File 281
6	Penryn 1 Box 11	11.6	Newspaper article. Regene	8.1.04	Newspaper piece	
7	Penryn 1 Box 11	11.7	Penryn Official Guide. Hisc	C1960's	Brochure	2 copies
8	Penryn 1 Box 11	11.8	Visit West Country holiday		Leaflet	
9	Penryn 1 Box 11	11.9	Glasney College pen and ir		Postcard	
10	Penryn 1 Box 11	11.10	Arrival of foreign cattle. Srr	24.12.1870	newspaper cutting	
11	Penryn 1 Box 11	11.11	Lady Jane Killigrew Festival	1999 & 1997	Brouchures	
12	Penryn 1 Box 11	11.12	Visit Historic Penryn. Histor		1 Sheet	
13	Penryn 1 Box 11	11.13	Cornish Theatre collective f		A5 Flyer	
14	Penryn 1 Box 11	11.14	History of a house. Cornwa		leaflet	
15	Penryn 1 Box 11	11.15	Street of Penryn. Sheets Lis		Printed A4s	
16	Penryn 1 Box 11	11.16	Letter from Penryn Town C	Year 1936	typed letter	

Next: files records and fields: graphic showing fields

- Tables consist of **columns** and **rows**
- Each row is a record (or Item)
- Columns are also known as **Fields**
- The Field represents the lowest level in the data hierarchy, it's where you input an individual piece of data

BF11 Penryn

Serial	BF Number	BF Serial No	Title/Description	Date/Period Covered	Document Type	Remarks
1	Penryn 1 Box 11	11.1	The story of the foundation	1845 - 1848	Typed A4 Document	2 copies
2	Penryn 1 Box 11	11.2	Jubilee of 150 years of Not	1845-1995	Souvenir Booklet	2 copies
3	Penryn 1 Box 11	11.3	Penryn versus Hilton Cleav	1953-1954	Souvenir Programme	
4	Penryn 1 Box 11	11.4	Seventy years of St.Glurias.	1908-1982	Document	
5	Penryn 1 Box 11	11.5	Tremough School. Small as	1976-1982	Brochures and Magazine	Blue File 28
6	Penryn 1 Box 11	11.6	Newspaper article. Regene	8.1.04	Newspaper piece	
7	Penryn 1 Box 11	11.7	Penryn Official Guide. Hisc	C1960's	Brochure	2 copies
8	Penryn 1 Box 11	11.8	Visit West Country holiday:		Leaflet	
9	Penryn 1 Box 11	11.9	Glasney College pen and ir		Postcard	
10	Penryn 1 Box 11	11.10	Arrival of foreign cattle. Srr	24.12.1870	newspaper cutting	
11	Penryn 1 Box 11	11.11	Lady Jane Killigrew Festival	1999 & 199	Brouchures	
12	Penryn 1 Box 11	11.12	Visit Historic Penryn. Histor		1 Sheet	
13	Penryn 1 Box 11	11.13	Cornish Theatre collective i		A5 Flyer	
14	Penryn 1 Box 11	11.14	History of a house. Cornwa		leaflet	
15	Penryn 1 Box 11	11.15	Street of Penryn. Sheets Lis		Printed A4s	
16	Penryn 1 Box 11	11.16	Letter from Penryn Town C	Year 1936	typed letter	

Serial	BF Number	BF Serial No	Title/Description	Date/Period Covered	Document Type	Remarks
1	Penryn 1 Box 11	11.1	The story of the foundation	1845 - 1848	Typed A4 Document	2 copies
2	Penryn 1 Box 11	11.2	Jubilee of 150 years of Not	1845-1995	Souvenir Booklet	2 copies
3	Penryn 1 Box 11	11.3	Penryn versus Hilton Cleav	1953-1954	Souvenir Programme	
4	Penryn 1 Box 11	11.4	Seventy years of St.Glurias.	1908-1982	Document	
5	Penryn 1 Box 11	11.5	Tremough School. Small as	1976-1982	Brochures and Magazine	Blue File 281
6	Penryn 1 Box 11	11.6	Newspaper article. Regene	8.1.04	Newspaper piece	2 copies
7	Penryn 1 Box 11	11.7	Penryn Official Guide. Hisc	C1960's	Brochure	
8	Penryn 1 Box 11	11.8	Visit West Country holiday:		Leaflet	
9	Penryn 1 Box 11	11.9	Glasney College pen and ir		Postcard	
10	Penryn 1 Box 11	11.10	Arrival of foreign cattle. Srr	24.12.1870	newspaper cutting	
11	Penryn 1 Box 11	11.11	Lady Jane Killigrew Festival	1999 & 1997	Brouchures	
12	Penryn 1 Box 11	11.12	Visit Historic Penryn. Histor		1 Sheet	
13	Penryn 1 Box 11	11.13	Cornish Theatre collective i		A5 Flyer	
14	Penryn 1 Box 11	11.14	History of a house. Cornwa		leaflet	
15	Penryn 1 Box 11	11.15	Street of Penryn. Sheets Lis		Printed A4s	
16	Penryn 1 Box 11	11.16	Letter from Penryn Town C	Year 1936	typed letter	

Next: fields have specific data types

Fields have specific data types

- There are different types of Field for different types of data

Data Types

- ArchiveDb currently supports the following data types:
 - Short text
 - Long text
 - Date
 - Date and time
 - Integer
 - Floating point number
 - Decimal
 - Geolocation
 - Binary

Short Text

- Up to 250 characters in length
- Indexed, searchable and sortable

Long text

- Any amount of text up to approximately 2 GB or about 1 billion characters
- Searchable but neither indexed nor sortable

Date

- Mandatory day, month and year, no time component
- Data entry validation
- Indexed, searchable and sortable
- Example: 23/04/2023

Date and time

- Mandatory day, month, year, hours, minutes, seconds components
- Data entry validation
- Indexed, searchable and sortable
- Example: 23/04/2023 15:37:56

Integer

- Whole numbers
- Signed 32-bit integer (no decimal places) -2,147,483,648 to 2,147,483,647
- Data entry validation
- Indexed, searchable and sortable
- Example: 42

Floating point number

- 3.4×10^{-38} and 3.4×10^{38} that can contain decimal values but is only precise to 7 decimal digits
- Data entry validation
- Indexed, searchable and sortable
- If you require greater precision, and particularly for financial or currency data, use the decimal data type
- Example: 42.73

Decimal

- Compound data type consisting of text prefix of up to 6 characters followed by a number up to 13 digits in length followed by 4 decimal places. Primarily intended to hold currency values accurately with up to 4 decimal places along with a currency prefix e.g. £4001.21.
- Spaces not allowed. Can be used for values other than currency that require a character prefix and a decimal number. Note you can only have the single decimal point otherwise you would need to have a short text field.
- Prefix and number components individually searchable
- Data entry validation
- Number component indexed and sortable
- Precise to 17 decimal digits

Geolocation

- Compound data type with latitude and longitude
- Data entry validation
- Searchable and indexed but not sortable
- Example: 50.1537526, -5.0681146

Binary

- Contains computer files such as images, PDFs, audiovisual files or any other computer file
- Natively supported binary attachment file types:
 - PDF
 - Image:
 - TIF, TIFF, JPEG, JPG, PNG, GIF, BMP
 - Multimedia:
 - MP3, MP4, WAV, MPG, MOV, WMV
 - HTML/MHT is partially supported
- Natively supported files are displayed by and manipulated by the user interface in ArchiveDb
- Other file types can be stored but will be opened externally in the associated application
- Binary files can either reside directly within the database itself or, by default, be stored on disk
- Multimedia files support a searchable timed transcript

Dates and datatypes

Gotchas with respect to date/datetime datatypes

- **Date/Datetime** data types require mandatory day, month, year and additionally, in the case of **Datetime**, hours minutes and seconds
- In historic contexts this level of mandatory precision is often inappropriate. Dates are often only known to a month and year or just the year or just a decade
- If you record an imprecise date in a **Date/Datetime** data type you still need to complete the mandatory day, month, year etc. components
- This leads to ambiguity as missing data is often entered as, for example the first day of the month and the first month of the year
- But did you really mean 1 January 1932?
- Or was it just the year 1932?

Dates and datatypes

Short text fields as a workaround

- The solution to imprecise dates is often to use a short text field

Pros:

- Instead of entering (eg) 1.1.1903 in a **Date** field, enter 1903 in a **Short Text** field and there is no ambiguity that the date is precise only to a year
- ArchiveDb has an inverted index of dates that allows dates entered in text fields to be found by searches
- When a record is saved the text data is parsed and any text that looks like a date is added to the inverted index

Cons

- Short text fields are not validated and are subject to typographical errors

Dates and datatypes: Part 2

The inverted index

- ArchiveDb has a special inverted index of dates that allows dates entered in text fields to be found by searches
- When a record is saved, the text in all text fields and attached text content (eg PDFs) is parsed, and any text that looks like a date is added to the inverted index. This includes numbers that may represent a year.
- The parsing process uses an algorithm to separate any text that looks like a date into year, month and day. Each of these components is stored in separate fields of the index. Text that looks like a date containing only a year, or year and month, will leave the missing components blank.
- This makes it possible for a date search to look for whole dates, or part dates such as year and month, or year alone

Inverted indices and parsing

- Parsing is the process of splitting a piece of text into its individual words
- The process of parsing looks at the properties of those words, for example if they might represent numeric or date data
- An inverted index is a data structure to store a mapping from content, such as words or dates in our case, to their locations in a database file or a document.
- This process allows fast retrieval times in a full text search
- ArchiveDb uses two inverted indices: one for words and one for dates

Schemas

- Schema is the name given to the field specification for a database entity such as a table
- It includes field names and captions, datatypes, field visibility and field order, along with other properties
- Schemas can be shared across multiple entities in ArchiveDb
- From ArchiveDb 3.0 all hierarchical elements (fonds, series, files etc) will be defined by user-designed schemas. This will allow users to flexibly utilise ISAD(G) field definitions for Fonds, Series, Files etc, fully, partially, or not at all

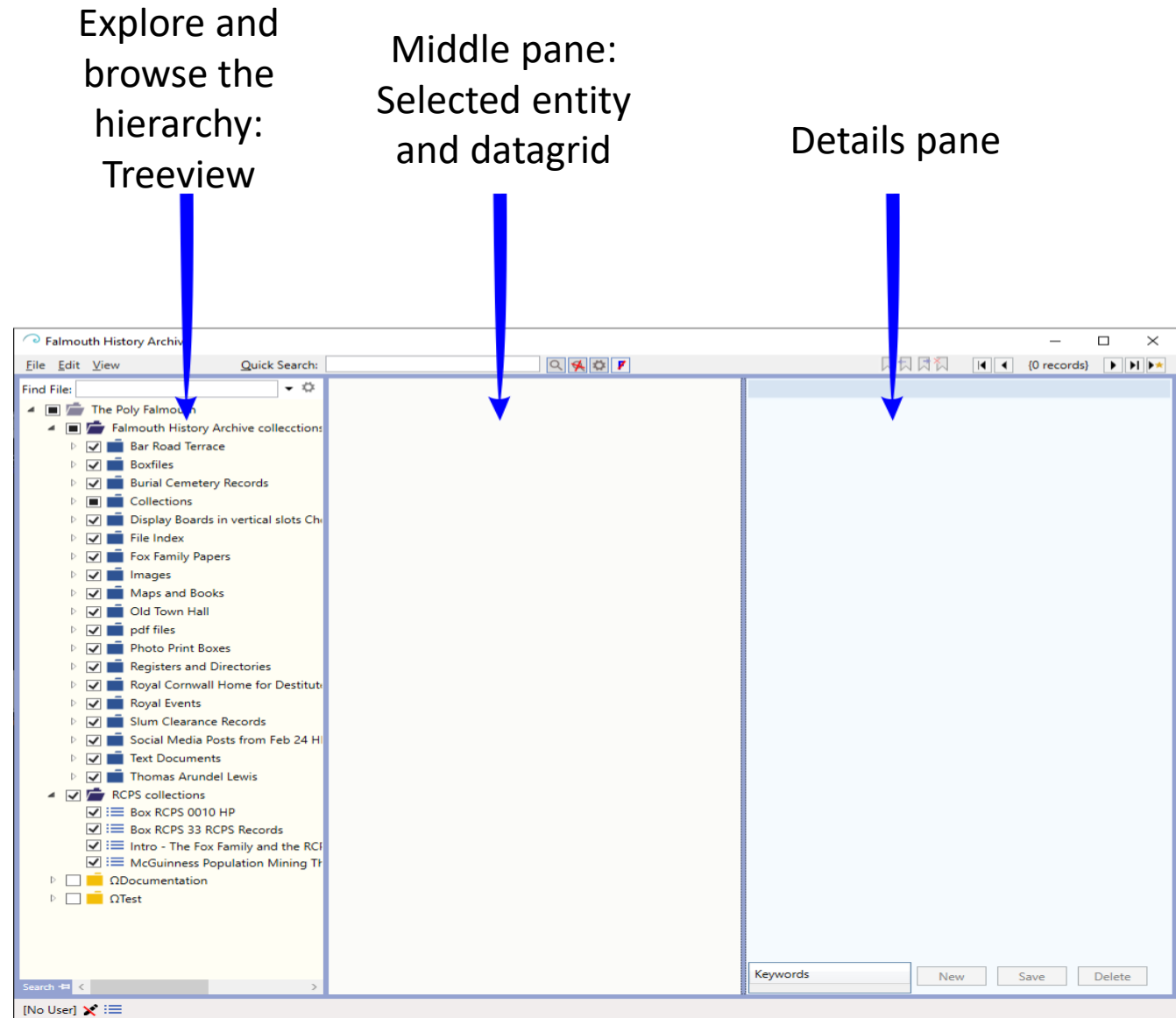
Part 2

The user interface

Remember to use the right click context menus

The user interface

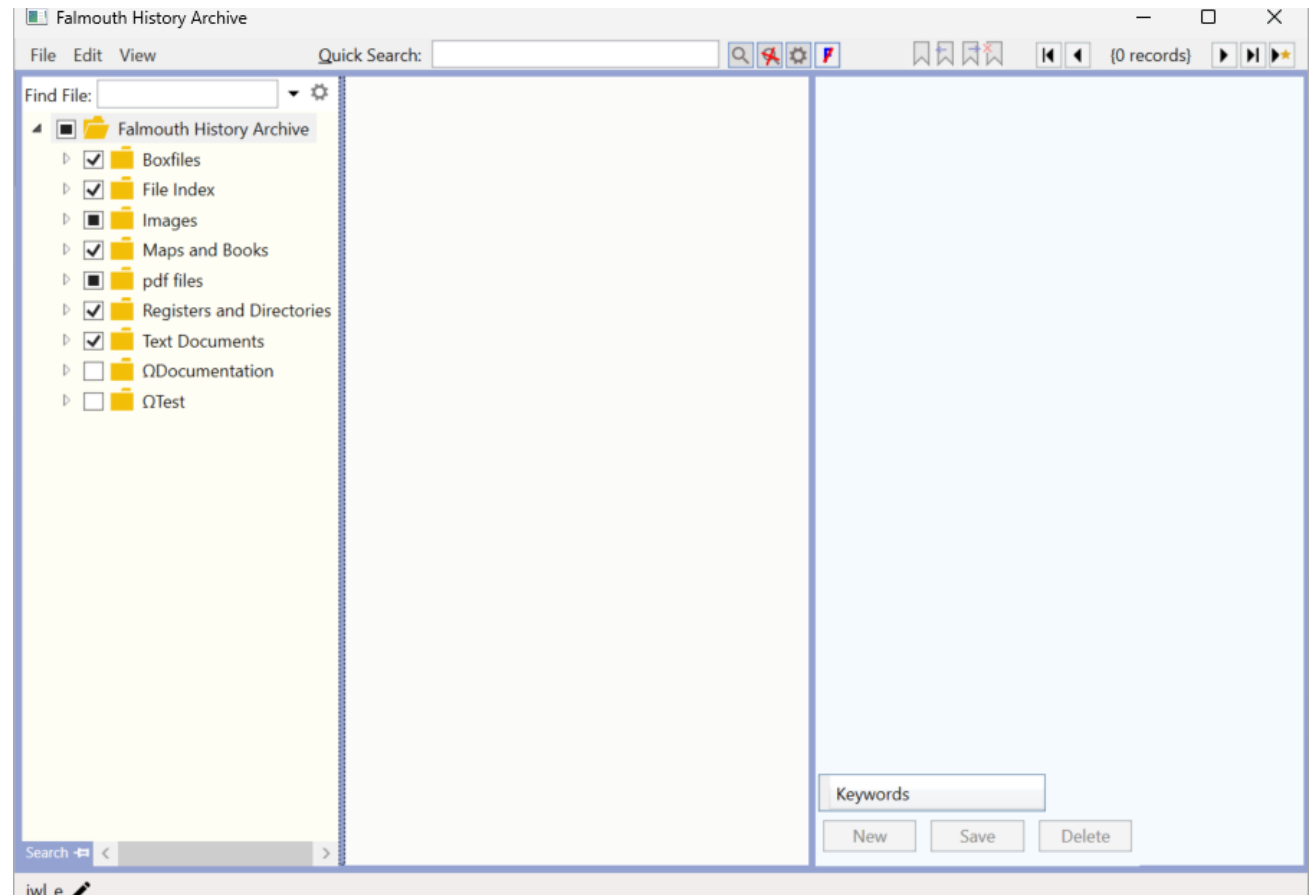
- The following is an overview of the user interface
- Full details are given in ArchiveDb documentation located in the Documentation file in the ΩDocumentation folder



Next: overview of main window features

Overview of main window features

- Quick search criteria box and buttons
- Bookmark buttons
- Record navigation buttons
- Advanced search button
- Menus
- 3 panes that drill down into detail from left to right
 - Data hierarchy Treeview
 - Datagrid panel
 - Data detail panel
- Resize panes by dragging dividers
- Hide panes with view menu commands



Next: Drilling down into data

Drilling down into data

Selected File

File records table
with selected
record

The screenshot shows the Falmouth History Archive application interface. On the left is a file tree under 'The Poly Falmouth' containing 'Falmouth History Archive collection', 'Bar Road Terrace', and 'Boxfiles'. The 'Boxfiles' folder is expanded, showing a list of files including '221010 Shelf 36 Box File', 'ACQ 573 BF12 subset File', 'BF10 Budock', 'BF100 Research Papers', 'BF10b Budock', 'BF11 Penryn', 'BF12 Flushing', 'BF13 Constantine & M...', 'BF14 Arwenack and the', 'BF15 Falmouth Docks', 'BF23 Working Boats Sail', 'BF24 Isolation Hospital', 'BF26 Log of HMS Amaze', 'BF27 Tourism', 'BF28 Arwenack and the', 'BF28A Arwenack House', 'BF29 Step Back in Time', 'BF3 Perranarworthal', 'BF34 Refuse Disposal', 'BF34a Refuse Destructor', 'BF35 Earles Retreat', 'BF4 Perran Foundry', 'BF46 Historic Homes of', 'BF47 Coastguard', 'BF5 Devoran and Redrut', 'BF50 Museum', 'BF52 Marlborough Hous', 'BF53 The Lizard', 'BF54 Railway Articles', and 'BF56 Local Historians CA'. The 'BF15 Falmouth Docks' file is selected. In the center is a table of file records with columns: Serial, BF Number, BF Serial No, and Title/Description. The table contains 34 records, with the 21st record, 'BF15 Falmouth Docks 15.7.4. Falmouth's New Dock', highlighted. On the right is a detailed view of the selected record, showing fields for Serial (21), BF Number (BF15 Falmouth Docks), BF Serial No (15.7.4), and Title/Description (Falmouth's New Dock – Quiet Opening Function – Heads of Shipping Industry Attend. Covers the inspection and inauguration of the 'new graving dock'; leading shipping magnate guests inc heads or personnel of Federal Steam Navigation Co, Anglo-Persian Oil Co, Cunard Co, etc; liner Hardwicke Grange first vessel to enter dock). Below these fields are 'Date/Period Covered' (September 1928), 'Document Type' (Photocopy of newspaper article), and 'Remarks' (1 s). At the bottom right are buttons for 'Keywords', 'New', 'Save', and 'Delete'.

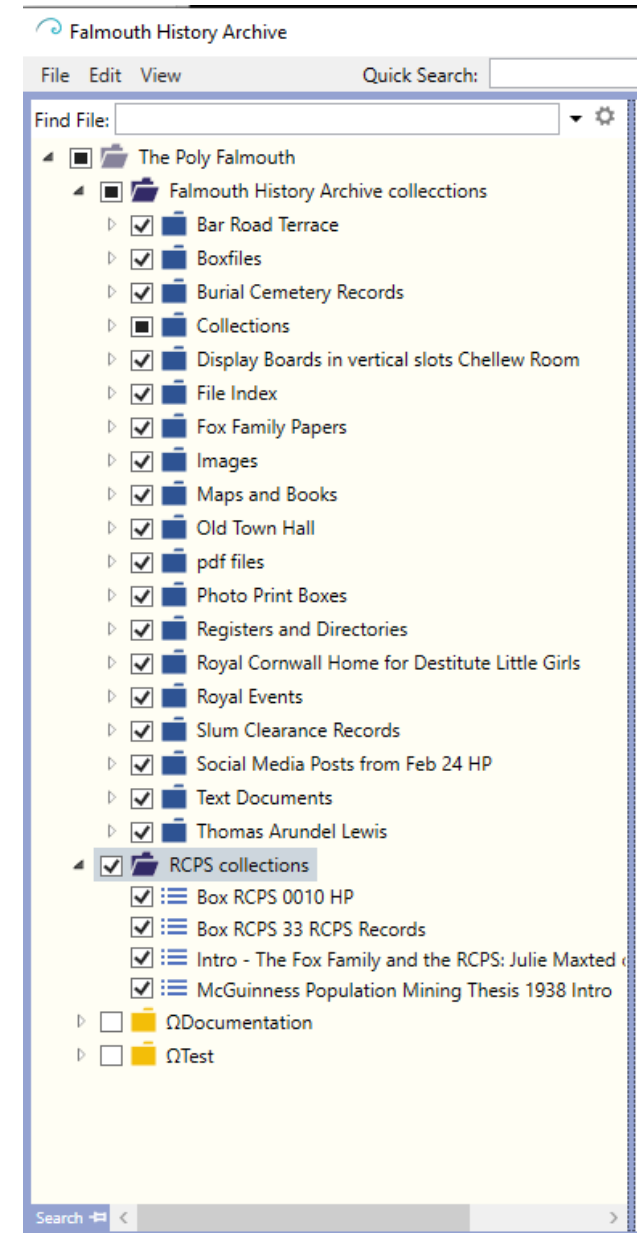
Serial	BF Number	BF Serial No	Title/Description
1	BF15 Falmouth Docks	15.1.1.	Award of BEM to
2	BF15 Falmouth Docks	15.1.2.	Progressive Falm
3	BF15 Falmouth Docks	15.1.3.	Falmouth Docks -
4	BF15 Falmouth Docks	15.1.4.	Company Renam
5	BF15 Falmouth Docks	15.2.1.1.	Report of the layi
6	BF15 Falmouth Docks	15.2.1.2.	Laying the Found
7	BF15 Falmouth Docks	15.2.2.	Early Years of the
8	BF15 Falmouth Docks	15.3	An Act for making
9	BF15 Falmouth Docks	15.4.1.	Falmouth Docks.
10	BF15 Falmouth Docks	15.4.2.	Brief Historical Ba
11	BF15 Falmouth Docks	15.4.3.	Map of Outer Har
12	BF15 Falmouth Docks	15.4.4.	Falmouth Yard La
13	BF15 Falmouth Docks	15.5.1	Ships Built at Falm
14	BF15 Falmouth Docks	15.5.2.	List of the vessels
15	BF15 Falmouth Docks	15.6.1.	Falmouth Harbou
16	BF15 Falmouth Docks	15.6.2.	The Falmouth Do
17	BF15 Falmouth Docks	15.6.3.	Falmouth Ship Re
18	BF15 Falmouth Docks	15.7.1.	Falmouth Shippin
19	BF15 Falmouth Docks	15.7.2.	Falmouth Docks
20	BF15 Falmouth Docks	15.7.3.	A Walk in the Dock
21	BF15 Falmouth Docks	15.7.4.	Falmouth's New Dock
22	BF15 Falmouth Docks	15.7.5.	The Foundry and
23	BF15 Falmouth Docks	15.7.6.	Early Years of Falm
24	BF15 Falmouth Docks	15.7.7.	Harbour of Falm
25	BF15 Falmouth Docks	15.7.8.	Photograph of th
26	BF15 Falmouth Docks	15.7.9.	BRITON WPD 20/
27	BF15 Falmouth Docks	15.7.10.	Laying the Found
28	BF15 Falmouth Docks	15.8	Falmouth Group
29	BF15 Falmouth Docks	15.9	Materials Receive
30	BF15 Falmouth Docks	15.10.	Description of Wk
31	BF15 Falmouth Docks	15.11.	Information Docu
32	BF15 Falmouth Docks	15.12.	Falmouth Docks
33	BF15 Falmouth Docks	15.13.	Falmouth Docks
34	BF15 Falmouth Docks	15.14	Sacrification Guir

Selected
record detail

Next: Main window – File Treeview

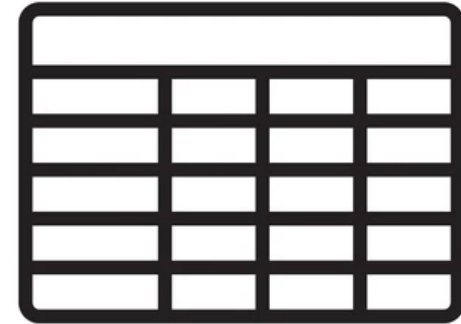
Main window – File Treeview

- The data hierarchy is central to ArchiveDb and is browsed via the file treeview
- Expand and collapse nodes
- 3 state hierarchy checkboxes
- 2 state File checkboxes
- Checked Files are included in searches, unchecked excluded
- Searching for file by name or keyword in **Find File**
- Browsing a file
- Treeview context menu
 - Reset checks
 - Add New file
 - Edit file
 - View file definition
- Checked vs Selected files

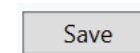
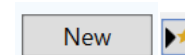


Main window – Datagrid pane – 1

- Tabulated data
 - Selected file
 - Search results
- Numeric fields are right aligned
- Text fields are left aligned
- Column headers can be clicked to sort the column
 - Long text and location do not sort, decimal values are sorted by the number component
- Data can be edited directly in the table
- You may briefly see 3 dots in a field ... this means data is being downloaded from the server
- New records need to be created explicitly with the New button
- Changes and new Records need to be explicitly saved



shutterstock.com · 1874648533



Main window – Datagrid pane – 2

- Above the data grid is **File name** and access rights icon
- Left of the File name is an expander button to view **File Details**



- Filename is a hyperlink if the Datagrid is a search result

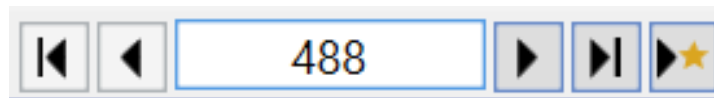


Record navigator

- Buttons for next, previous, first, last and new record
- Current record number and count





- Type into text box to navigate to specific record number

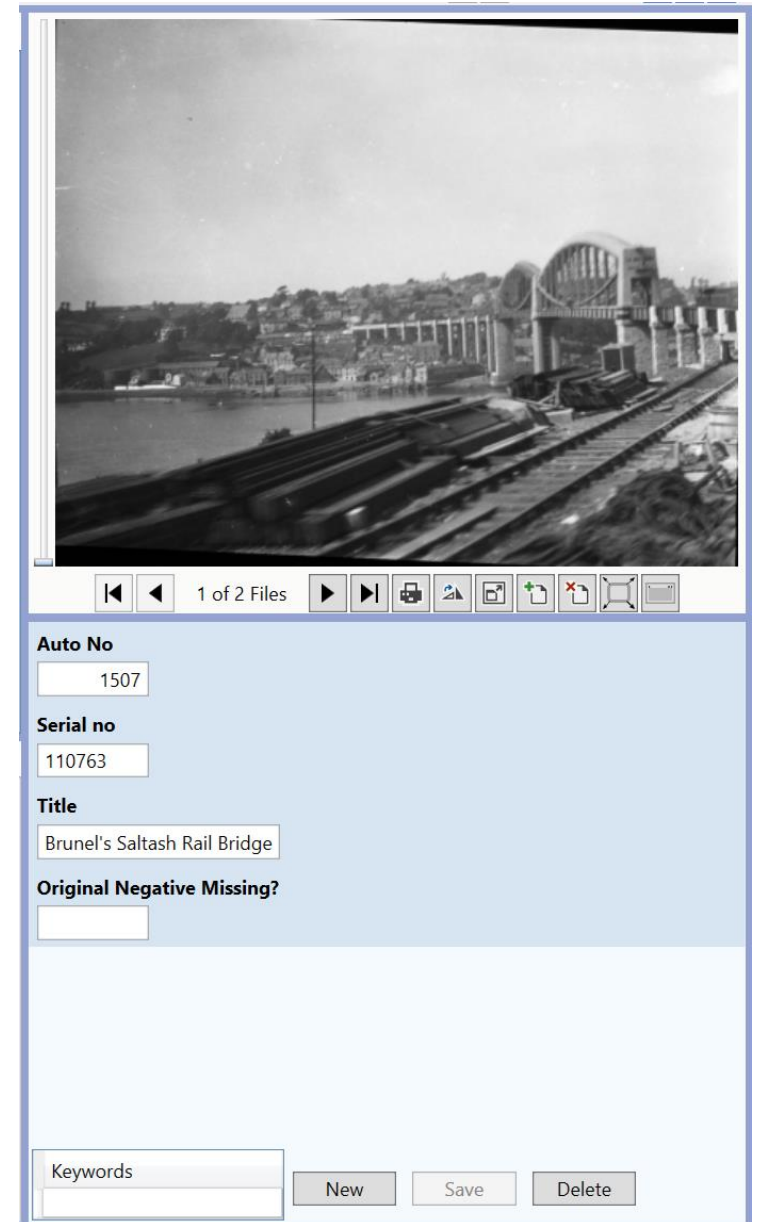


Main window – Details pane

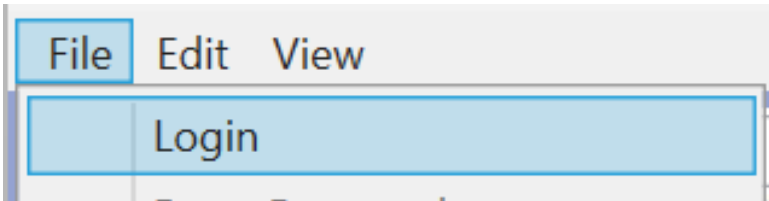
- Binary files (media) appear at the top with a draggable divider
- Text data is left-aligned, numeric data is right-aligned
- Add data directly to text boxes and use the Tab key to navigate
- Validation errors are indicated by a red border and an ! exclamation mark
- Tooltips show the data type of each text box
- Context menus are available on fields and the panel
- To add location data, right-click a point on Google Maps, copy the latitude and longitude, and paste it into the location field using context menu
- Location data can also be dragged from Google maps
- Keywords can be added to the table, bottom left, or via the field context menu
- New, Save and Delete buttons

Binary data pane

- Displays Media
- Common function buttons at bottom of pane
- Context menus on PDFs and images
- Zooming and panning
- Fullscreen and separate window with always-on-top button
- Save image as...
- Export image as...
- Support for timed transcripts on audiovisual files 
- Other file types open in own applications 

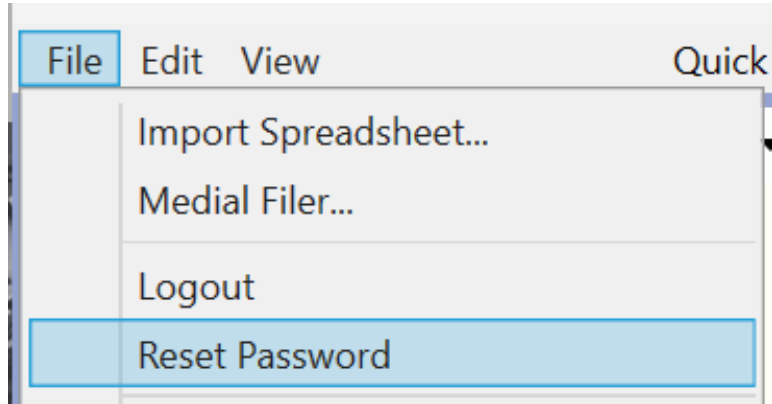


Users and login



- ArchiveDb opens in read-only mode to provide easy public access
- Users should log in
 - Logged in users see all files and can add, edit and delete and have their own bookmarks
- Users are added by administrators
- User rights can be:
 - Read only
 - Editors
 - Administrators
- Editors can add edit and delete records, add and edit files
- Administrators can delete files and change the hierarchy and perform additional administrative tasks

Users and login



- Once you have been given username by a system administrator, login to ArchiveDb and change your password
- ArchiveDb accepts passwords of minimum 6 characters but with no special character requirement
- ArchiveDb complies with industry standard for safe storage of passwords

Next:

Adding and editing Files

- Open the File Edit dialog from the Treeview context menu
- Edit file dialog has 3 tabs
 - File description
 - Schema
 - Permissions

The screenshot shows a 'File Edit' dialog box with three tabs: 'File Description', 'Schema', and 'Permissions'. The 'File Description' tab is active. It contains the following fields:

- File Title:** A text box containing 'Documentation'.
- File Code:** A text box containing '1'.
- Description:** A large text area.
- Notes:** A large text area.
- File Keywords:** A section with a header 'Keywords' and a large text area below it.
- Search by default:** A checkbox that is currently unchecked.
- ISAD(G) Fields:** A section with a dropdown arrow and a large text area.
- Add Media:** A button with a plus sign icon and the text 'Add Media'.

At the bottom right of the dialog are 'OK' and 'Cancel' buttons.

Next: description of what's on the schema tab

Schema tab

- System fields at the top
- User-defined fields at the bottom
- Note the function of **Hide in Grid View** and **Never Show**
- It is possible to change the field order
- Do not change data type or field name once records have been added to a file
- More details are to be found in the documentation

The screenshot displays the 'Schema tab' interface. At the top, there are two sections: 'Select New or Existing:' with radio buttons for 'New' (disabled) and 'Existing' (selected), and 'Select Existing Schema:' with a dropdown menu showing 'BF10BUDOCK'. Below this is a 'Schema Name:' field containing 'BF10BUDOCK'.

The main area is divided into two sections: 'System Fields' and 'User Defined Fields'.

System Fields:

Field Name	Data Type	Hide in Grid View	Never Show	Description	Field C
SYSTEMDATETIME	Date + Time	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		0
ID	Integer (64 bit)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		1
FK_FILES	Integer (64 bit)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		2
ID_PARENT	Integer (64 bit)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		3

User Defined Fields:

Field Name	Caption	Data Type	Hide in Grid View	Never Show	Description	Field Ord
SERIAL	Serial	Integer (32 bit)	<input type="checkbox"/>	<input type="checkbox"/>		5
BFNUMBER	BF Number	Short Text	<input type="checkbox"/>	<input type="checkbox"/>		6
BFSERIALNO	BF Serial No	Short Text	<input type="checkbox"/>	<input type="checkbox"/>		7
TITLEDESCRIPTION	Title/Description	Long Text	<input type="checkbox"/>	<input type="checkbox"/>		8
DATEPERIODCOVERE	Date/Period Covered	Short Text	<input type="checkbox"/>	<input type="checkbox"/>		9
DOCUMENTTYPE	Document Type	Short Text	<input type="checkbox"/>	<input type="checkbox"/>		10
REMARKS	Remarks	Short Text	<input type="checkbox"/>	<input type="checkbox"/>		11

At the bottom right, there are 'OK' and 'Cancel' buttons. At the very bottom, it says 'Record count: 172 File ID: 559'.

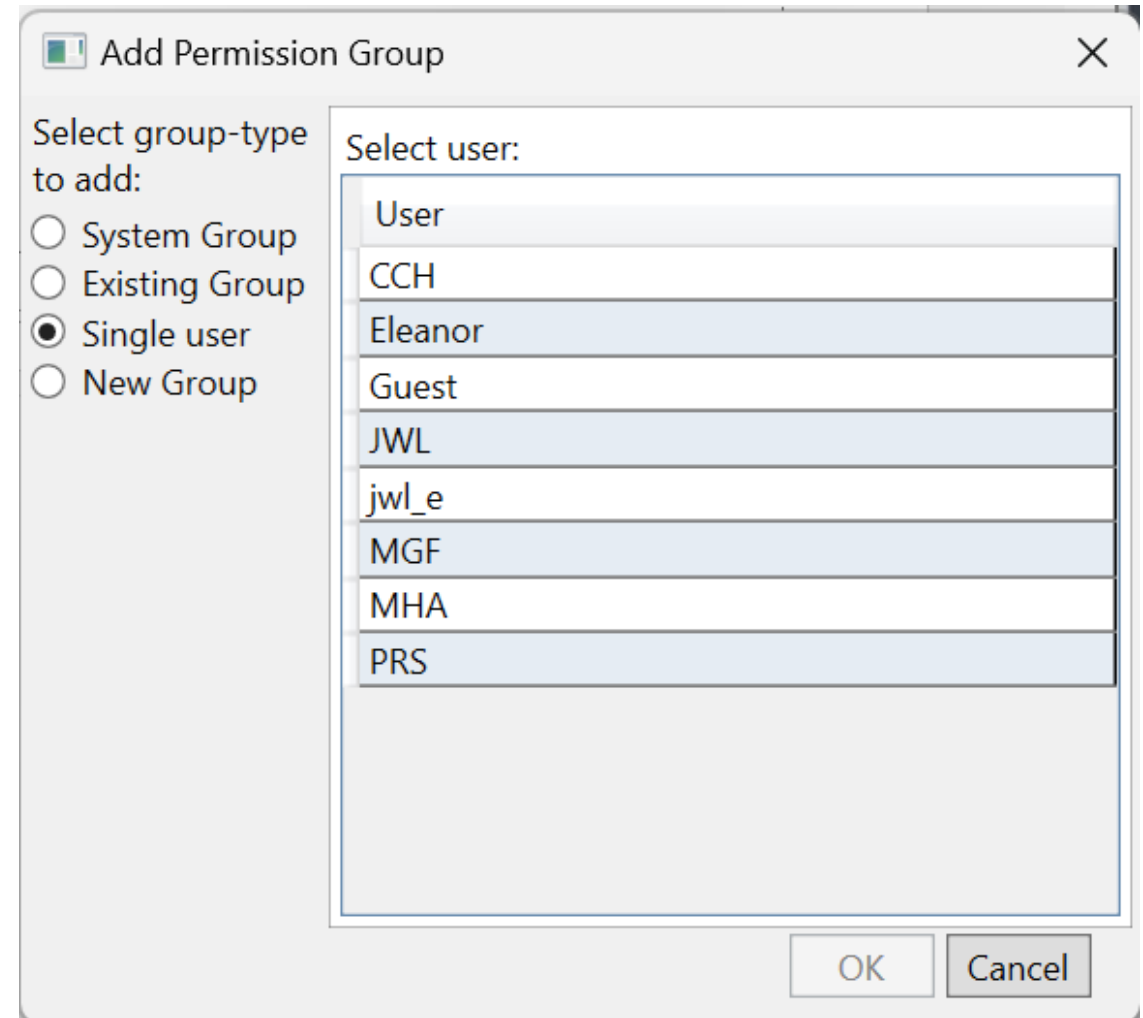
Next: permissions tab

Permissions tab

- By default, files have unrestricted access and are visible without requiring a login
- Restrict permission by adding group of users.
- There are 3 types of group
 - System
 - User-defined
 - Single user
- System groups automatically contain users defined by their login permission,
- System ***All Users Group*** is everyone but excludes viewing without login.

Permission group types

- System
 - Administrators
 - Editors
 - All users – includes administrators, editors and read-only users
- User-defined groups (on the existing groups radio button)
- Single user groups
- Use the System/All Users group to restrict data access to logged-in users



Next: main window – description of data grid pane

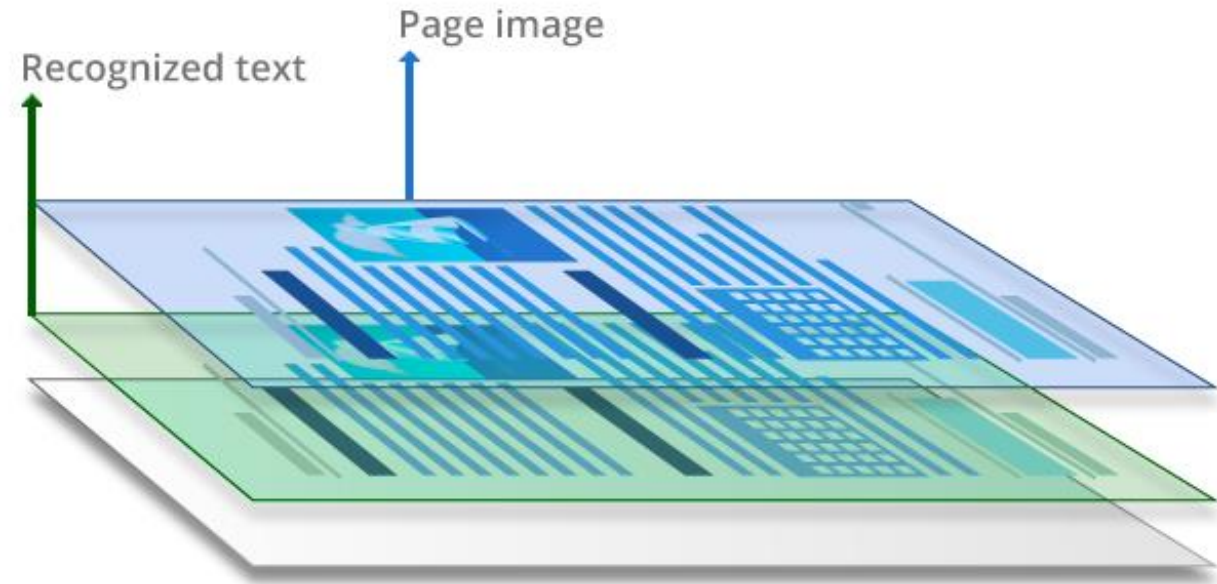
The PDF is ArchiveDb's core searchable document

Types of PDF file ¹

- True (Digitally Created) PDFs:
 - True PDFs are created using software like Microsoft Word, Excel, or by printing from an application (using a virtual printer). They contain both text and images. Characters in the text and meta-information have electronic character designations. You can easily search, select, edit, or delete text in true PDFs.
- Image-Only (Scanned) PDFs:
 - When you scan hard copy documents or convert images (JPEG, TIFF, screenshots) into PDFs, you create image-only PDFs. These PDFs “lock” the content in a snapshot-like image. They lack an underlying text layer, making them non-searchable. You cannot modify or mark up the text in image-only PDFs. However, you can make them searchable by applying OCR (Optical Character Recognition) to add a text layer beneath the page image.
- Searchable [image] PDFs:
 - Searchable PDFs result from applying OCR to scanned PDFs or other image-based documents. During OCR, characters and document structure are analyzed and “read.” A text layer is added to the image layer, making these PDFs fully searchable. You can select, copy, and mark up text in searchable PDFs. Text recognition can occur via scanning devices, desktop OCR software, mobile apps, web-based services.

Searchable PDFs

- Layers in searchable PDF
- Graphic copied from ABBYY FineReader



Next: ISO standard PDFs

PDF file types continued ¹

ISO standard types in bold

- **PDF (Standard PDF):**
 - Most common PDF format is the. It encapsulates a fixed-layout document containing text, fonts, vector graphics, and links. This format is suitable for general document distribution.
- **PDF/A (Archival PDF):**
 - PDF/A is designed for long-term file storage and archiving. It restricts certain features (such as JavaScript, audio, and video content) to ensure compatibility over time. Archivists and managers often use this format.
- **PDF/E (Engineering PDF):**
 - PDF/E supports construction, engineering, and manufacturing specifications. It's commonly used in those industries for sharing technical information.
- **PDF/X (Printing PDF):**
 - PDF/X is optimized for graphic designers and print professionals. It better supports graphics, ensuring accurate printing results.
- **PDF/VT (Variable Data Printing PDF):**
 - Similar to PDF/X, PDF/VT is used by print professionals and graphic designers. It includes additional customization features for variable data printing.
- **PDF/UA (Universal Accessibility PDF):**
 - PDF/UA enhances readability and navigation for people with disabilities. It's compatible with assistive technology.
- PAdEs (PDF Advanced Electronic Signatures):
 - PAdEs sets standards for advanced electronic signatures, aligning with major legislation.
- PDF Healthcare:
 - Developed to secure best practices for handling and managing healthcare information.
- **Searchable PDF:**
 - A searchable PDF is essentially a standard PDF file with a search function. It makes image-based PDFs text-searchable.

PDF/A

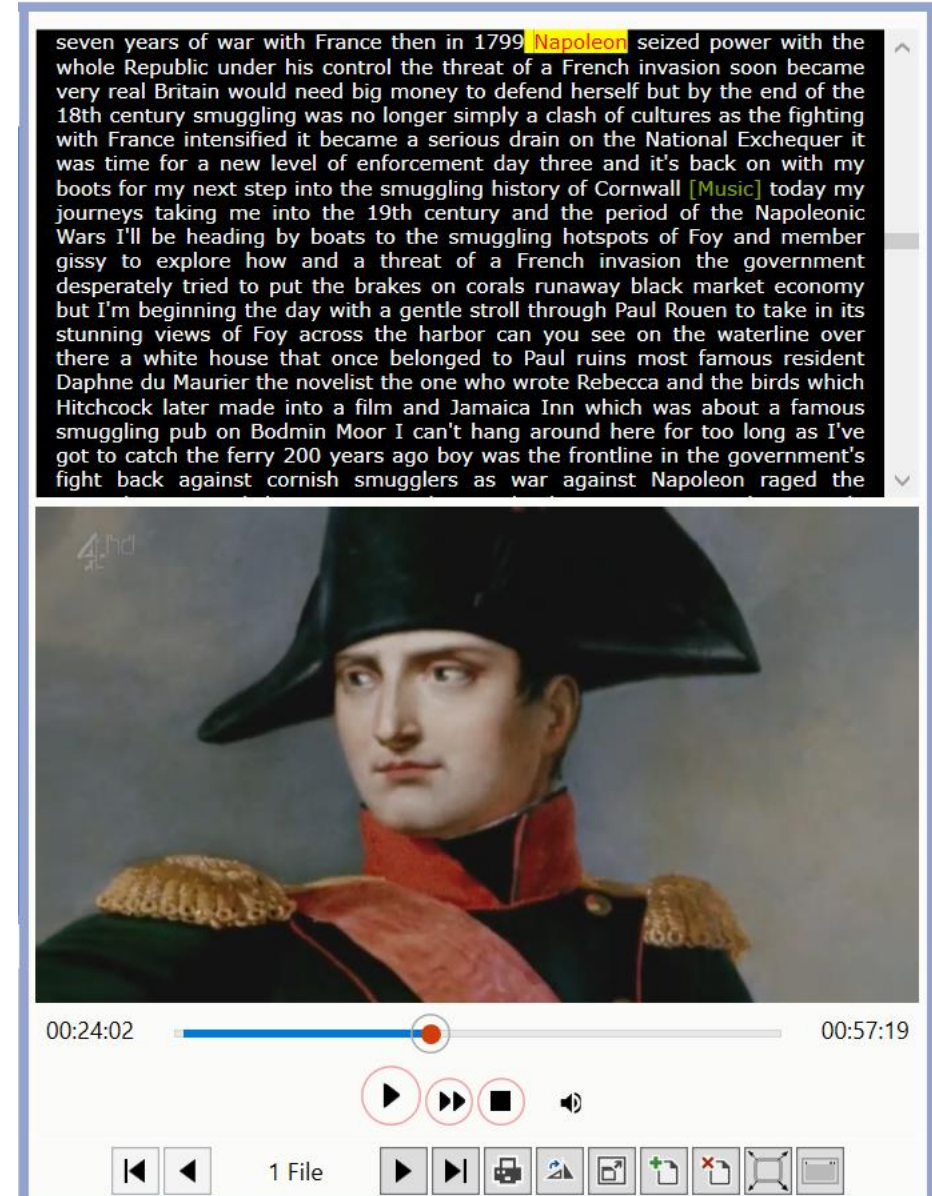
- PDF/A is designed for long-term file storage and archiving. It restricts certain features (such as JavaScript, audio, and video content) to ensure compatibility over time.
- This option is available when generating PDFs from Word documents and searchable PDFs from OmniPage. Ensure the option is selected
- ISO standard

Why use PDFs and not Word documents?

- ArchiveDb displays and manipulates PDF files natively
- ArchiveDb can provide a full text search of PDF files, which are indexed on import
- ArchiveDb can highlight text in a PDF search
- PDF/A is ISO standard for long-term storage
- Word documents can only be displayed in their own application and sit outside ArchiveDb when imported
- Word document formats not ISO standardised
- ArchiveDb has no access to Word document content and can neither search nor highlight text within a Word document
- Display of Word documents is dependent upon having Word, or a Word compatible program, installed on the workstation hosting ArchiveDb

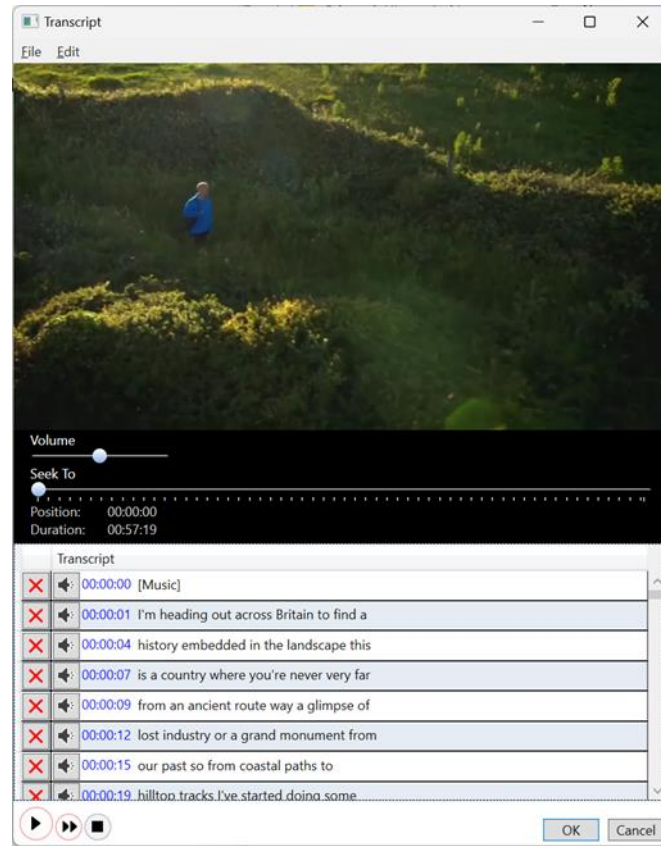
Audiovisual files

- Searchable timed transcripts
- Oral histories



Next: reminder graphic for timed transcripts

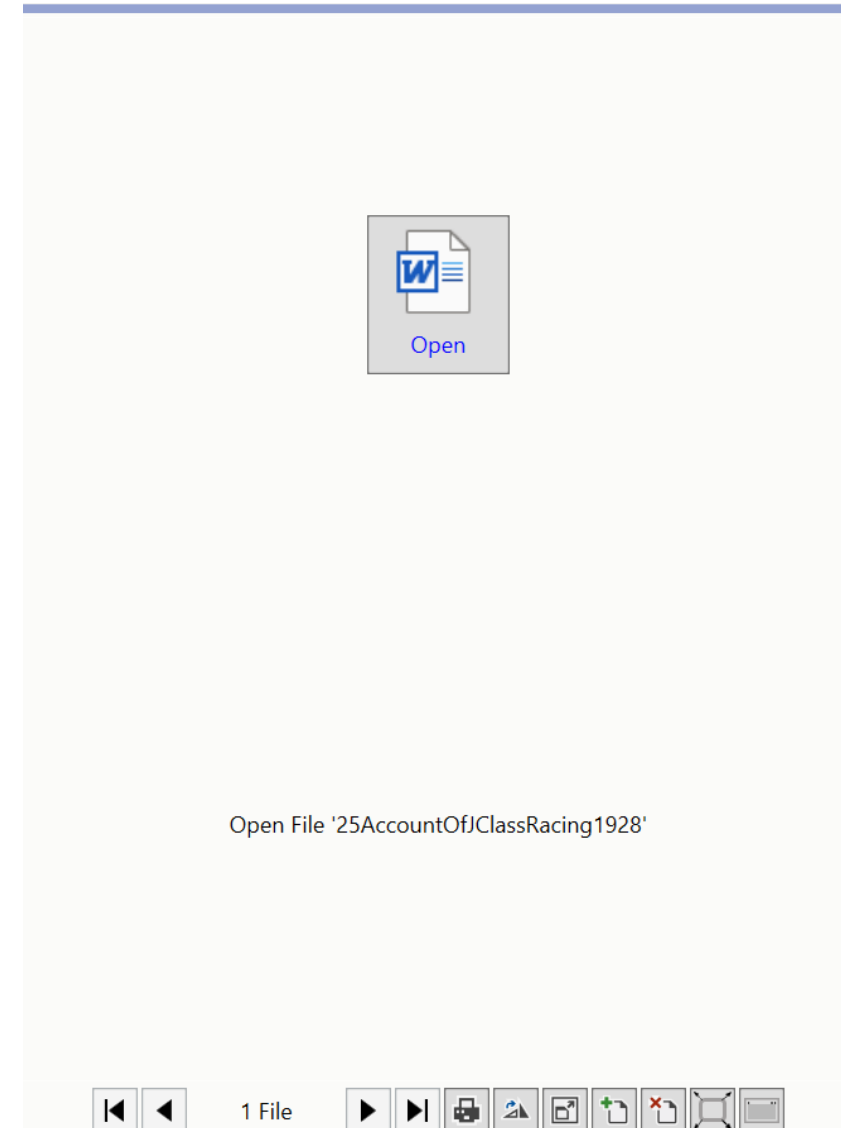
Audiovisual files – timed transcripts




Next: display other file types

Other file types

- Displayed as an icon that opens the file in its own application
- Requires the application to open the file to be installed on the workstation hosting ArchiveDb
- There is no restriction on the types of file that can be attached to an ArchiveDb record



Bookmarks

- Record place markers 
- Bookmark records that you wish to come back to, e.g. found in a search or records you are working on
- Toggle on and off
- Belong to the current user
- Cannot be shared
- Persist between sessions if logged in, otherwise they are temporary
- Bookmark is shown top right in the field section of the details pane and in the record header (extreme left) of a row in the data grid

Part 3

Ways to add data and annotations

Next: description of the 4 ways to add data

Ways to add data

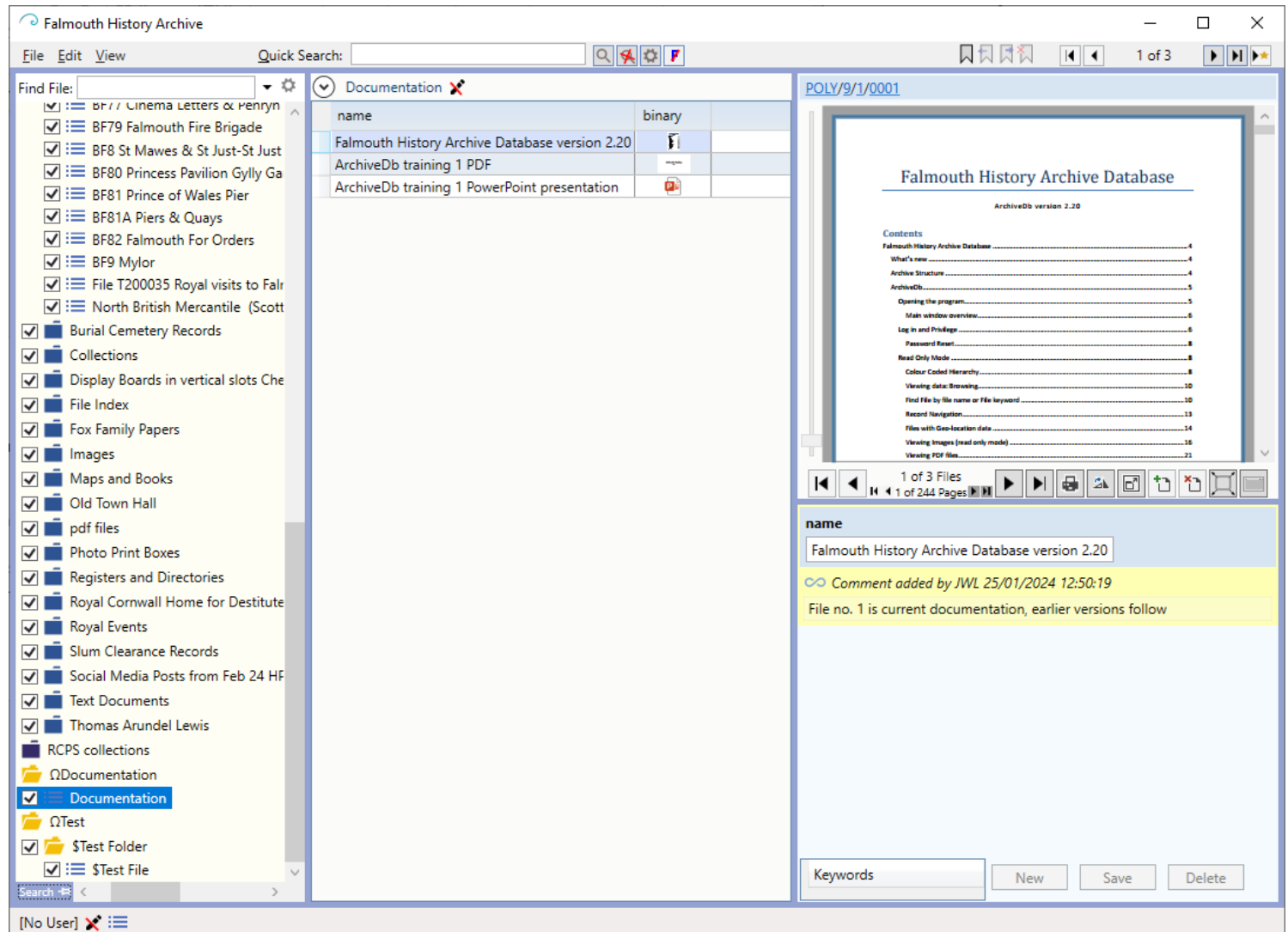
- Directly to tables - directly in the data grid or record detail pane
 - New records must be explicitly created using either the new button on the record navigator or the new button on the record detail pane
 - All changes must be explicitly saved
- Spreadsheet import
 - Create new File or append to existing
- Media Filer
 - Streamlined bulk-add Images, PDFs or other media to new records
- Ad hoc data
 - You can add any additional piece of data to any record. Most useful for adding images and locations to record who schema does not include the appropriate data type. Supports drag-and-drop and pasting

Annotations

- Annotations sit outside the data record and don't form part of the archive data
- Searchable with Detailed Search
- You can set the annotation scope: everyone, just you, or a specific group of users
- Use for any purpose and scope accordingly
- User groups are shared with file permission groups, take care if deleting

Annotations

- Annotations are yellow, reminiscent of post-it notes, and placed in the details pane. They consist of a caption and annotation text.
- The icon to the left of the caption indicates scoping [all users, group or individual user]



Part 4

Searching

Searching

- 'Quick search' (QS)
- Detailed or advanced searches
 - Selected file
 - Single schema
 - Multi-schema (checked file search)

‘Quick search’ (QS)

- Word and date search of whole database (checked files)
- Utilises the word and date inverted indices created when a record is saved
- Includes the content of attached PDFs and audiovisual transcripts
- Significant words only
- Includes dates within narrative text and PDF content
- File keywords and record keywords
- Single criteria box at the top of the main window
- Algebraic syntax



Quick search (QS)

- Quick search is not Google
- Google supports semantic matching and natural language searching
- QS is a word and date matching search

Detailed or advanced searches

- Comprehensive searching of data fields in record data
- Excludes the content of attachments
- Includes numeric and other data
- Microsoft Access-like criteria grid for entering complex criteria

Quick search (QS)

- Whole words, no wild cards (available in detail/advanced searches)
- Fuzzy option for spellings and plurals
- Criteria must include an indexed word
- Results are in a 3-column table 
- Search options 
- Quick search from text context menu
- Results are read-only

QS Results – 3 column table

Quick Search:

- Clickable headers to sort results
- 1st column is the name of the file containing the record
- Middle column is a nested list. Click an item to highlight its location in the record, PDF or AV file
- 3rd is a précis of the record's data
- At the top of the results grid is a hyperlink to the selected record with an expander showing the file detail
- Filter buttons for media content
- Grouping of results by File

All 16 Text 3 Image 1 PDF 11 AV 1

QS Criteria

- 'AND' and 'OR' criteria
- With **AND**, it's like saying, "Show me records that meet ALL these conditions."
- With **OR**, it's "Show me records that match ANY of these conditions."
- Multiple words in the criteria box are implicitly 'AND'ED'
- Search string – string is the name given to the text you type

QS AND criteria

- Search the database for 'James'



1980

- Search the database for 'Bull'



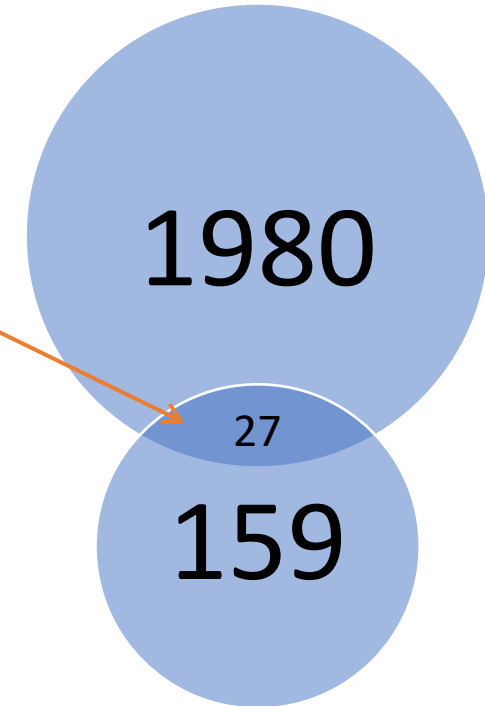
159

QS 'AND' criteria continued

- Search the database for 'James Bull'
- Search returns 27 records that contain both words, 'James' and 'Bull'
- The search string 'James Bull' contains an implicit 'AND' and is the equivalent of typing:

James Bull

James + Bull



QS 'OR' criteria

- Specify '**OR**' criteria by using '/'
- Search returns the 2105 records containing either 'James' or 'Bull', including those that contain both
- $1980 + 159 - 34 = 2105^{\S}$
- *§ ArchiveDb excluded some records where 'James' and 'Bull' were too far apart to qualify as a match; there are 34 records containing both names*

James / Bull



2105

QS Combining 'AND' and 'OR' criteria: algebraic syntax

- If you wish to search the database for '*workhouse*', '*workhouse*' may be spelt as a single word or as 2 words, '*work house*'
- To search for both forms of '*workhouse*' you can combine an 'AND' criteria with an 'OR' criteria
- The **AND** criteria is: '*work + house*'
- To combine this with an **OR** criteria, use parentheses for the **AND/OR** sets: '*(work + house)/workhouse*'
- Parentheses can be omitted from single word sets, as shown above

Algebraic syntax: adding levels

- To find 'union workhouse' or 'union work house', use algebraic syntax
- Start with the previous search string:
'(work + house)/workhouse'
- Enclose the whole thing within parentheses to create a new level:
'((work + house)/workhouse)'
- Add 'union' as an **AND** criterion: *'union + ((work + house)/workhouse)'*
- Add as many levels as you need but, beware, it becomes confusing

Quick search (QS) – Date searching

- The indexing process indexes possible dates or partial dates found within the content text of PDF documents, audiovisual transcripts, short and long text fields and date fields.
- Enclose date criteria between two exclamation points.
- Within the exclamation points you can add either just the year, a month plus year, or day, month, year
- Tolerant of date formats in both indexing data and criteria input
- You must use a 4 digit year.
- Date criteria can be included within the algebraic syntax

QS some example dates to try

- !1917!
- !jan 1917!
- !16 jan 1917!
- !17 jan 1917!
- !feb 1917! (15,17,2,28)
- !1915!
- !jan 1915!
- !jan 1923,
- 1832

Quick search (QS) – Keyword searching

- Precede record keyword criteria with #
- Precede file criteria with @
- Can be included within the algebraic syntax

Quick search (QS) – Exact string matching

- Enclose criteria within double quotes ""
e.g. *“richard fox”*
- Can be included within the algebraic syntax

Quick search (QS) – Fuzzy search

- Finds similar sounding words, plurals and some spelling errors, e.g. *chasewater railway* also finds *chacewater railway*
- May take a long time

QS: Search options – basic tab

- Fuzzy on/off
- Result sorting options
 - Natural order
 - Average matches
 - Total match score
 - Filename: ascending, descending
- Group results by file name
- All partial matches




QS: Search options – advanced tab

- Match qualifier distance – the maximum number of words between matches
- Edit distance and Jarowinkler
- Fine tuning fuzzy search criteria

Detailed (advanced) searches

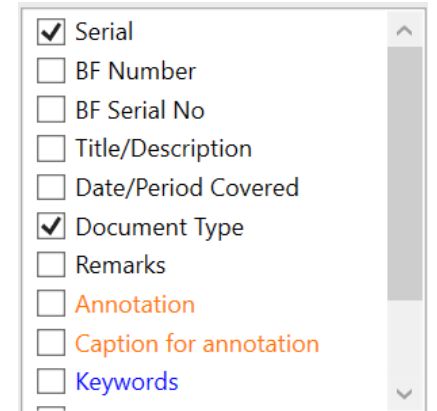
- 3 flavours:
 1. Individual files (Selected File search)
 2. Files sharing the same table design i.e. Shared Schema Search
 3. Files with different schema (Checked File Search)
- Hidden by default, available from the **View** menu or **Search** button bottom left of Treeview
- Query grid reminiscent of Microsoft Access

Selected File search

- Searches individual, named, fields including Annotations and Keywords
- Check field names required for criteria or sorting 
- Use the query grid to apply criteria to more than one field 
- Run button to execute the search 
- Results can be edited

The query grid – selected file/shared schema

- Check field names to add fields to the query grid
- Apply criteria to more than one field
- Specify result sort order using the **Sort** combo-boxes



A vertical list of fields with checkboxes. The checked fields are 'Serial' and 'Document Type'. Other fields include 'BF Number', 'BF Serial No', 'Title/Description', 'Date/Period Covered', 'Remarks', 'Annotation' (in orange), 'Caption for annotation' (in orange), and 'Keywords' (in blue). The list has a scrollbar on the right.

Field Name	Serial	Document Type
Source Table	Integer (32 bit)	Short Text
Sort	NotSorted ▾	NotSorted ▾
Criteria		
OR		

Next: criteria within the query grid

Criteria in the query grid

- Add criteria to the cells in the grid
- Criteria along a row are AND criteria
- Criteria on different rows are OR criteria

Field Name	Serial	Document Type
Source Table	Integer (32 bit)	Short Text
Sort	NotSorted ▾	NotSorted ▾
Criteria	>5	like "**newspaper**"
OR	<5	like "**a4**"

Syntax

- Basic syntax: use a comparison operator followed by criterion
- Numeric criteria are straightforward, use a comparison operator followed by the number criterion (e.g. >100)
- To match specific text values, enclose them in double quotes (“”)
- To match specific Date values, enclose them between hashes (##)
- Use the **Like** keyword for wildcard searches
- Combine criteria within a cell with **AND** and **OR**

Comparison operators

- =, >, <, >=, <= equal to, greater than, less than, greater than or equal to, less than or equal to
- Between *val1* And *val2*
- is null, is not null: specifies empty field or a field containing a value
- NOT (field does not contain specified value) equivalent to <>,
- Val1 **AND** val2, val1 **OR** val2
- **Like** and **Not Like** for wildcard searches




Wildcards

- Wildcards flexibly search for patterns within text data
- Use with **Like** or **Not Like** operators
- Use * or % to represent zero or more characters in a string
- Use ? or _ to represent exactly one character in a string
- Like “a*” matches *Annie, Amelia, Alice* etc
- Like “*newspaper*” finds *newspaper* or *newspapers* anywhere in a field
- Like “wh?te” matches *White* and *Whyte*

Shared Schema Search

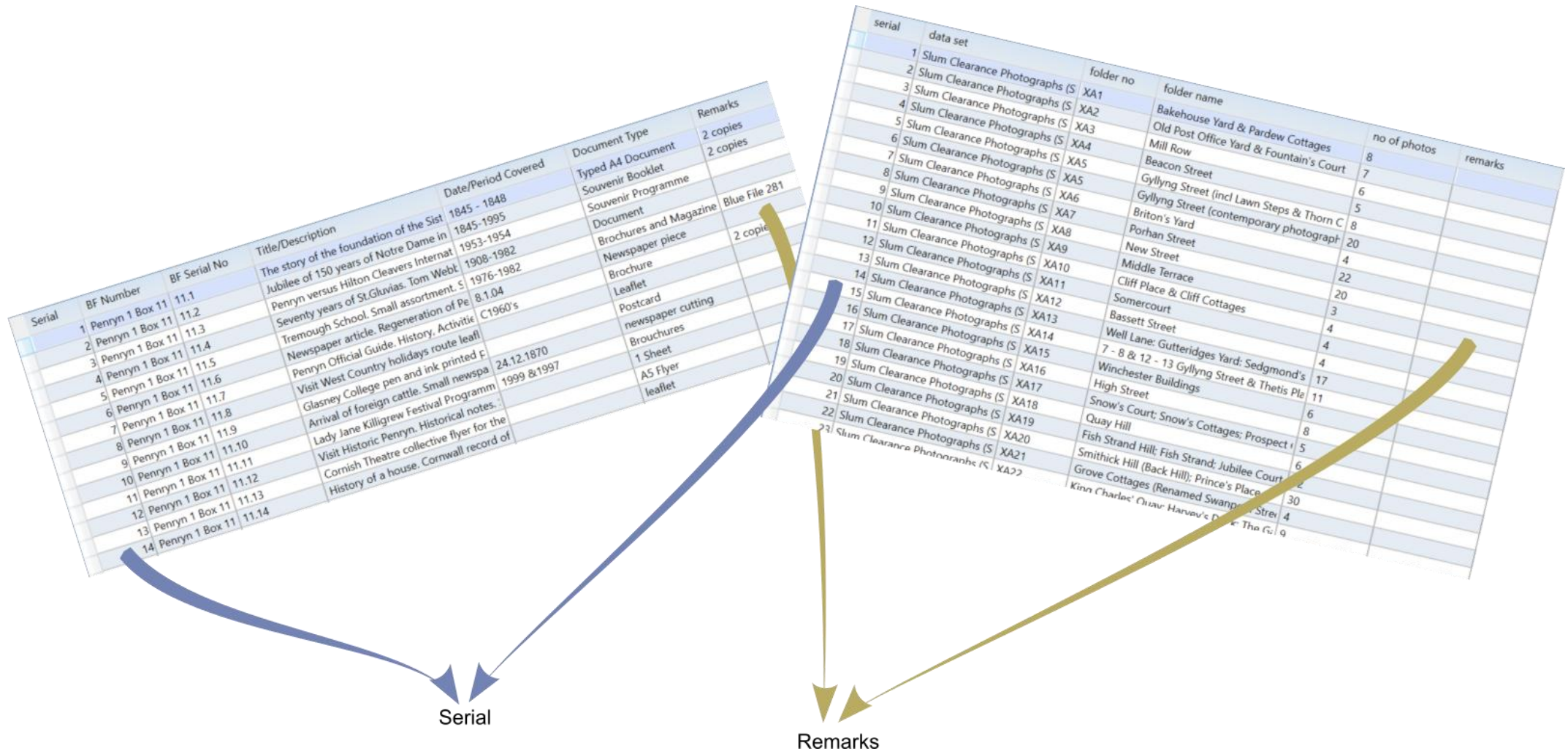
- As for the selected file search but includes all files using the same schema (table design)
- Files sharing a schema can be selected/deselected for search purposes, as required
- Changes the checked files in Treeview
- Option to revert to previous check state on Selected File search
- File name hyperlink displays file name for the selected record
- Results contain a File ID column
- Results can be edited

Checked Files Search

- Searches checked files whatever their schema
- Search is primarily by datatype
- Common/shared fields can be searched by name 
- Results show only the columns selected in the search 
- Datagrid contains more than 1 result per record 
- Results can be edited in detail pane, not datagrid

Checked Files Search: Dissimilar Schemas - a

- 2 common fields



Next: Checked Files Search – shared fields

Checked Files Search – shared fields

- Common/shared fields within the checked files **can** be searched by name
- All other fields **must** be searched [more generically] by Datatype
- *41aSlum Clearance Photographs - XA Folders + BF11 Penryn* share **Serial & Remarks**
- Datagrid column header is fieldname for shared field criteria
- For datatype criteria the datagrid column header displays the datatype, while each cell shows the field name in gray followed by its value

Checked Files Search – shared fields

- Example *41aSlum Clearance Photographs - XA Folders + BF11 Penryn* checked
- *Remarks* and *Serial* are shared fields and listed beneath their datatypes
- *Remarks* and *Serial* have been checked to add them to the query grid
- ***Between 7 and 10*** has been set as the criteria for *Serial*
- No criteria is set for *Remarks*
- Click **Run** to run the search...

The screenshot shows a software interface for searching through checked files. At the top, there are two tabs: 'Selected File/Schema' and 'Checked Files'. Below the tabs, there are several groups of radio buttons for selecting fields and datatypes. The 'Remarks' and 'Serial' fields are checked. Below these, there are more options like 'Geo Location', 'Keywords', 'My Annotations', and 'Primary Record'. At the bottom, there is a table for defining the search criteria.

Field Name	Serial	Remarks
Source Table	Integer (32 bit)	Short Text
Sort	NotSorted	NotSorted
Criteria	between 7 and 10	
OR		

At the bottom right, there are 'Cancel' and 'Run' buttons.

Checked Files Search – shared fields: output

- Only the fields in the query grid are output
- Shared (named) fields have field names as column headers
- File ID is always shown
- Results can be sorted by File ID – click the header
- Note example use of the **Between** operator

The screenshot displays the 'Checked Files Search' interface. At the top, a 'Find File:' search bar contains the text 'xa'. Below it, a list of files is shown, with '41aSlum Clearance Photographs - XA Folders' selected. To the right, a table displays the search results:

File ID	Serial	Remarks
88	7	
88	8	
88	9	
88	10	
96	7	2 copies
96	8	
96	9	
96	10	

Below the results table, the 'Selected File/Schema' tab is active, showing a 'Checked Files' section. This section contains several radio button options for field types: Short Text, Long Text, Integer, Date, DateTime, Number, and Decimal. The 'Remarks' field is checked under Short Text, and the 'Serial' field is checked under Integer. Below these are options for Geo Location, Keywords, My Annotations (with sub-options for Annotation text and Caption), and Primary Record (with sub-options for ID, SYSTEMDATETIME, and LASTMODIFIED). At the bottom, a 'Field Name' table is visible:

Field Name	Serial	Remarks
Source Table	Integer (32 bit)	Short Text
Sort	NotSorted	NotSorted
Criteria	between 7 and 10	
OR		

The interface concludes with 'Cancel' and 'Run' buttons at the bottom right.

Checked Files Search – shared/unshared: output

- *Short text* checked instead of *Remarks*
- Multiple rows per record: 1 per short text field
- Shared field *Serial* has fieldname as header
- Datatype *Short Text* had datatype header with fieldname in grey and data in black in data cell
- Rows belonging to the same record may be separated if a sort order is applied

The screenshot displays the 'Checked Files Search' interface. At the top, a 'Find File:' search bar contains the text 'xa'. Below it, a list of files is shown, with '41aSlum Clearance Photographs - XA F' selected. To the right, a preview table shows the contents of the selected file. The table has three columns: 'File ID', 'Serial', and 'Short Text'. The data is as follows:

File ID	Serial	Short Text
88	7	DATASET: Slum Clearance Photographs (Slum Clearan
88	7	FOLDERNO: XA6
88	7	FOLDERNAME: Briton's Yard
88	7	NOOFPHOTOS: 4
88	7	REMARKS:
88	8	DATASET: Slum Clearance Photographs (Slum Clearan
88	8	FOLDERNO: XA7

Below the preview table, there are several tabs: 'Selected File/Schema', 'Checked Files', and 'Search'. The 'Checked Files' tab is active. It contains a series of radio buttons for selecting the datatype: 'Short Text' (selected), 'Long Text', 'Integer', 'Date', 'DateTime', 'Number', and 'Decimal'. There are also checkboxes for 'Remarks' and 'Serial'. Below these, there are more options: 'Geo Location', 'Keywords', 'My Annotations' (with checkboxes for 'Annotation text' and 'Caption'), and 'Primary Record' (with checkboxes for 'ID', 'SYSTEMDATETIME', and 'LASTMODIFIED'). At the bottom, there is a table with the following structure:

Field Name	Serial	[Short Text]
Source Table	Integer (32 bit)	Short Text
Sort	NotSorted	NotSorted
Criteria	between 7 and 10	
OR		

Next: Checked Files Search - Keywords

Checked Files Search - Keywords

- You can add *Keywords* to the criteria grid by checking its radio button
- In this example **Is Not Null** is used as the criterion to find any records in the checked files having a keyword

The screenshot displays the 'Checked Files Search' interface. At the top, the 'Find File' search bar contains 'xa'. Below it, a list of files is shown with checkboxes and radio buttons. The 'Keywords' radio button is selected. To the right, a table shows the results of the search, with columns for 'File ID' and 'Keywords'. The table contains two rows: one with '96' and 'GLUVIAS', and another with '70' and 'PDF'. Below the file list, the 'Selected File/Schema' tab is active, showing a grid of search criteria. The 'Keywords' radio button is selected, and the 'Criteria' dropdown is set to 'is not null'. The 'OR' button is also visible.

File ID	Keywords
96	GLUVIAS
70	PDF

Field Name	Keywords
Source Table	Data keywords
Sort	NotSorted
Criteria	is not null
OR	

Next: Checked Files Search - Annotations

Checked Files Search - Annotations

- Annotation text and captions can be searched independently
- The search results will include only annotations accessible to you

Find File: xa

Documentation

File ID	Annotation text	Caption
525	Current documentation is file no. 1	Comment added by JWL 23/01/2

Selected File/Schema: Checked Files

Short Text Long Text Integer Date DateTime Number Decimal

Geo Location Keywords My Annotations Primary Record

My Annotations: ☒ Annotation text ☒ Caption

Primary Record: ☐ ID ☐ SYSTEMDATETIME ☐ LASTMODIFIED

Field Name	Annotation text	Caption
Source Table	Annotations	Annotations
Sort	NotSorted	NotSorted
Criteria	is not null	
OR		

Next: Checked Files Search – Primary record fields

Checked Files Search – Primary record fields

- *ID*, *SYSTEMDATETIME* and *LASTMODIFIED* fields from the Primary Record can be included in queries
- *SYSTEMDATETIME* and *LASTMODIFIED* are *DateTime* fields and so criteria require a time component – or use **Between Date1 and Date2** where Date1 and Date2 are Dates
- In this example, records in the checked files are sorted in descending order based on their last modification date (*LASTMODIFIED*), with the most recent ones appearing at the top

The screenshot displays the 'Checked Files Search' interface. At the top, a 'Find File:' search bar contains 'xa'. Below it, a list of files is shown with checkboxes and expand/collapse icons. The '74 Falmouth Borough Deeds Box' is selected and expanded, showing a list of files with their IDs and last modified dates. The 'Selected File/Schema' tab is active, showing the 'Checked Files' section. This section includes a grid of search criteria buttons: Short Text (selected), Long Text, Integer, Date, DateTime, Number, Decimal, Geo Location, Keywords, My Annotations (with sub-options for Annotation text and Caption), and Primary Record (with sub-options for ID, SYSTEMDATETIME, and LASTMODIFIED). The 'Primary Record' section is expanded, showing the selected fields. Below the criteria buttons is a table with the following structure:

Field Name	ID	LASTMODIFIED	[Short Text]
Source Table	DATA	DATA	Short Text
Sort	NotSorted	Descending	NotSorted
Criteria			
OR			