

Topic	Description
File Extension	mdb, accdb
MIME Type	None / <a href="#">application/octet-stream</a>
Structure	Proprietary relational database
Versions	Access 1.0 - 2003, 2007 - 2016
Primary fields or areas of use	Storing relational data. Used in many disciplines. Surveys, data entry, geospatial, storing metadata for certain disciplines, etc.
Source and affiliation	Microsoft
Metadata standards	No widely adopted standards. Built-in documentation may be used
Key questions for curation review	<ul style="list-style-type: none"> <li>• Intended use: use as a db or just be able to view the data?</li> <li>• Does the data exist in other formats?</li> <li>• Are there “linked” tables?</li> <li>• Is there documentation about each table and relationship?</li> </ul>
Tools for curation review	<ul style="list-style-type: none"> <li>• A Windows PC with Access (Access does not run on Mac). Recent versions of Access (e.g., 2016) should be able to open down to Access 2000</li> </ul>
Date Created	2019-01-18
Created by	Fernando Rios (University of Arizona Libraries, Office of Digital Innovation & Stewardship, <a href="mailto:frios.email.arizona.edu">frios.email.arizona.edu</a> ), Dave Fearon (Johns Hopkins University Libraries Data Services, <a href="mailto:dfearon@jhu.edu">dfearon@jhu.edu</a> )

Suggested Citation: Fernando Rios and Dave Fearon. (2019). Microsoft Access Data Curation Primer. Retrieved from the University of Minnesota Digital Conservancy. <http://hdl.handle.net/11299/202810>.

*This work was created as part of the Data Curation Network “Specialized Data Curation” Workshop #1 co-located with the Digital Library Federation (DLF) Forum 2018 in Las Vegas, Nevada on October 17-18, 2018. See also: Primers authored by the workshop attendees at DLF. <http://datacurationnetwork.org>.*

Date updated and summary of changes made	2019-02-01 - 1.0: initial version 2019-04-18 - 2.0: DCN peer review
--	--

Note: This primer assumes a conceptual familiarity with relational databases (and associated terminology) and a basic level of experience with Microsoft Access. For introductory material on relational database concepts and MS Access see

- [Relational Databases and Microsoft Access by Ron McFayden](#)
- [Intro to SQL: Querying and managing data from Khan Academy](#)
- [SQL for Librarians from Library Carpentry](#)

## Description of Format

A Microsoft Access (MS Access) database could be in one of two main formats: MDB and ACCDB. There are variations on these<sup>1</sup> but they will not be discussed further. The MDB format is the native format for MS Access 1.0 through MS Access 2003. Note that although the MDB file extension was used for all those versions, the internal format changed significantly. Starting with Access 2007, the ACCDB format was used by default.

The following descriptions of the MDB and ACCDB formats are largely summarized from the Library of Congress format description entries<sup>2</sup> along with additional information on the differences between the formats. Refer to those for more detailed information of the format along with a large list of additional resources.

	MDB	ACCDB
Overview	<ul style="list-style-type: none"> <li>• No public specification</li> <li>• Better compatibility with more versions of Access</li> <li>• User-level security</li> <li>• Database replication supported</li> <li>• Based on the JET database engine</li> <li>• Many well-developed 3rd-party tools</li> </ul>	<ul style="list-style-type: none"> <li>• No public specification</li> <li>• Better integrations with Outlook and web services</li> <li>• Can attach binary data to tables</li> <li>• Improved encryption</li> <li>• Based on the ACE database engine</li> <li>• Fewer 3rd-party tool support</li> </ul>
Versions	<p>The default format up until Access 2003</p> <ul style="list-style-type: none"> <li>• Access 1.0, Support was dropped in Microsoft Access 2010.</li> <li>• Access 2.0. Support was dropped in Microsoft Access 2010.</li> </ul>	<p>The default format starting from Access 2007</p> <ul style="list-style-type: none"> <li>• There are subtypes of the format but they have not been described</li> </ul>

<sup>1</sup> [https://en.wikipedia.org/wiki/Microsoft\\_Access#File\\_extensions](https://en.wikipedia.org/wiki/Microsoft_Access#File_extensions)

<sup>2</sup>MDB: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000462.shtml>

ACCDB: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000463.shtml>

	<ul style="list-style-type: none"> <li>• Access 95, aka Access 7.0. Support was dropped in Microsoft Access 2013.</li> <li>• Access 97. Support was dropped in Microsoft Access 2013.</li> <li>• Access 2000. Maximum size increased from 1GB to 2GB.</li> <li>• Access 2002-2003. New features, including PivotTable views, support for XML, and an updated version of SQL would result in files not compatible with Access 2000.</li> </ul>	<ul style="list-style-type: none"> <li>• Microsoft documentation indicates that using newer versions of Access (e.g., Access 2016) will result in databases that cannot be opened in earlier versions of Access.</li> </ul>
Metadata	<ul style="list-style-type: none"> <li>• No widely adopted standard, Access has its own way of storing the documentation of fields, tables, etc.</li> <li>• The SAIRD tool has some preservation metadata. See the Preservation section below</li> </ul>	

## Examples

### Common use cases

- Tabular data associated with spatial information - geodatabases
- Capturing survey results
- Storing metadata about other datasets
- Data entry forms - Makes use of the way Access can be used to easily create forms for more efficient data entry.
- Basic statistical analysis functionality and other data analysis uses such as content coding.

### Dataset Examples

These are drawn from a variety of disciplines and contain entries which are and are not associated with publications. Like many other kinds of datasets, there is a large variability in the amount of documentation included. The examples from TDAR are perhaps the best documented as they follow a structured approach.

- [https://figshare.com/articles/MS\\_Access\\_database\\_Seed\\_data\\_from\\_rattlesnake\\_GI\\_tracts/6034904](https://figshare.com/articles/MS_Access_database_Seed_data_from_rattlesnake_GI_tracts/6034904)
- [https://figshare.com/articles/EPA\\_Literature\\_Database\\_Cleft\\_Palate\\_Chemical\\_Data/6062773](https://figshare.com/articles/EPA_Literature_Database_Cleft_Palate_Chemical_Data/6062773)
  - Content linked to [newftp.epa.gov/COMPTOX/Virtual\\_Tissues\\_Data/EPALitDB/LitDB\\_projectsANDpublications/cleftpalate/cp\\_chems.accdb](http://newftp.epa.gov/COMPTOX/Virtual_Tissues_Data/EPALitDB/LitDB_projectsANDpublications/cleftpalate/cp_chems.accdb)
- [https://figshare.com/articles/Matura\\_NP\\_ESA\\_vegetation\\_survey/5630335](https://figshare.com/articles/Matura_NP_ESA_vegetation_survey/5630335)
- [https://figshare.com/articles/Picoplankton\\_cytometry\\_accdb/3180280](https://figshare.com/articles/Picoplankton_cytometry_accdb/3180280)
- [https://figshare.com/articles/Murray\\_Cod\\_Larval\\_Database\\_Murrumbidgee\\_2011-2013/5715097](https://figshare.com/articles/Murray_Cod_Larval_Database_Murrumbidgee_2011-2013/5715097)
- [https://figshare.com/articles/Effects\\_of\\_Recreation\\_on\\_Animals\\_Revealed\\_as\\_Widespread\\_through\\_a\\_Global\\_Systematic\\_Review/4297430](https://figshare.com/articles/Effects_of_Recreation_on_Animals_Revealed_as_Widespread_through_a_Global_Systematic_Review/4297430)
- <https://core.tdar.org/geospatial/394771/bell-site-geodatabase-mdb-file>
- [https://core.tdar.org/dataset/396397/gfs2011\\_nabone-database](https://core.tdar.org/dataset/396397/gfs2011_nabone-database) (locked)
- <https://core.tdar.org/dataset/394059/akurvik-access-database>

- Lasda Bergman, Elaine, 2014, "Replication data for: Citations", <https://doi.org/10.7910/DVN/27655>, Harvard Dataverse, V1
- Correlates of War, 2007, "National Material Capabilities (v3.02)", <https://hdl.handle.net/1902.1/10170>, Harvard Dataverse, V1, UNF:3:B4DsVV/TykpPeJQNXdAA== [fileUNF]

## Viewing

Recent version of Access (e.g., Access 2016) can still open and save databases down to Access 2000 however functionality may be changed when migrating from an older to newer or newer to older format. There is no version for Macintosh OS.

For opening very old Access versions, the Library of Congress provides resources:

- As of early 2017, Microsoft still provides [instructions online for importing an Access 2.0 database into versions of Access from 2000 through 2007](#).
- As of early 2017, Microsoft still provides [instructions online for importing an Access 95 database into Access 2007](#).

A basic requirement of opening these very old formats is a copy of Access 2007. If it is anticipated that there could be the need to open and curate these old databases, it is recommended to procure and maintain a virtual machine with a working copy of Access 2007 to allow migrations between older and newer versions of Access.

## Key questions to ask yourself

- Is documentation about the columns included? I.e., a codebook. The codebook is especially critical to understand the relationships between the tables
- Are the primary/foreign key relationships described? Exporting individual tables to a spreadsheet format does not keep these relationships so it's critical to document them (at least include an exported version of the Relationships diagram alongside the exported tables.)
- How easily could this database be exported into flat CSVs? The answer will depend on the above points
  - If exporting is easy i.e., if there are no relationships between tables and there is no embedded code or forms that are integral to the purpose of sharing the dataset. Suggest simply exporting it rather than archiving the database itself, since the database adds no particular functionality.
- Is the data contained in the database sourced from somewhere else?
  - It could be that the database was used as a way to organize other data.
  - If the data is available elsewhere, it might be that it's in a more FAIR format there and it would be worth pointing that out in the metadata
- What version of Access was used?
  - Most of the time, this won't be stated so you'll just have to try to open it and see if it works
- Is your version of MS Access secure?
  - Make sure options to disable macros and other code from executing automatically is disabled since you don't know if the dataset could be malicious. The default of newer versions of Access is to disable such functionality

## Key clarifications to get from researcher

### Very useful clarifications

- For data entry databases with forms, a description of the data entry workflow would help with understanding the underlying database structure
  - What is the main entry form? Where does the entry workflow start? Etc.
- Documentation of the meaning of the relationships in the database, not just which relationships exist (the latter can simply be obtained via the Relationship Diagram which is auto-generated by Access)
- Is the data available natively in a non-Access format elsewhere?

### Nice-to-have clarifications

- Documentation description of database table fields, ideally within the database design metadata.
- Documentation of macros and modules, ideally as comments within the code.
- What version of Access was used?
- Which tables and/or queries do specific results from the paper come from (in cases where the database is specifically associated with a publication)

## Applicable metadata standard, core elements and readme requirements

**Accessing metadata schema:** Most database platforms have standard means of exporting database schema: descriptions of the database structure and design, including the tables, fields, relationships, and, sometimes the calculations or procedures used by the database. In Access, the Database Documenter (on the Database Tools tab, Fig. 1) is the direct means of generating reports on the database schema.

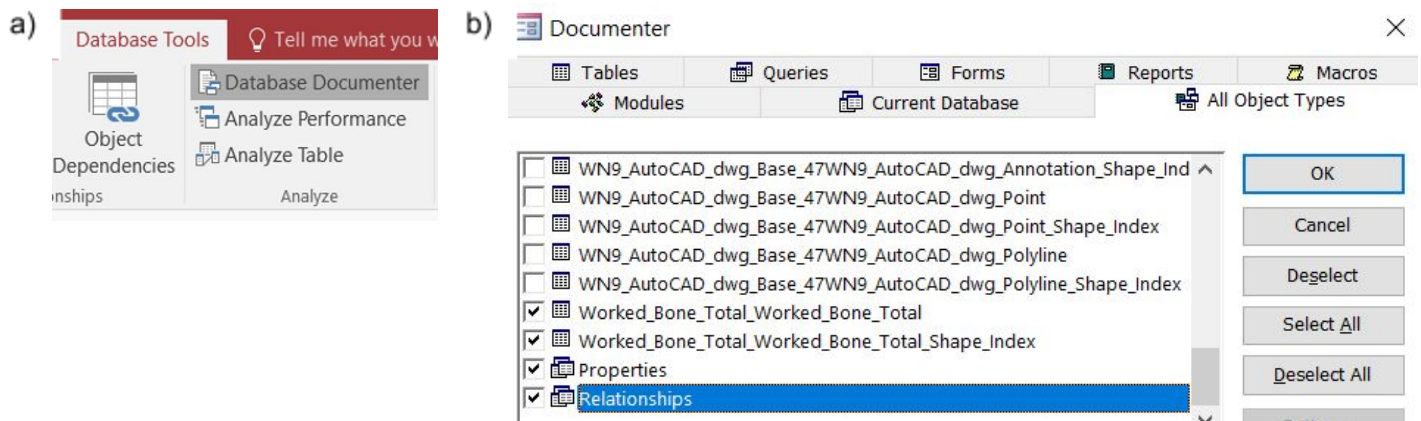


Figure 1: Database Documenter

The Documenter report includes a lengthy property report of each field, including description. Exporting as .txt format is recommended for utility. The decision to generate schema reports for curation may depend on plans for preservation rather than documentation that is of direct use to researchers interested in the data itself, since much of the content consists of properties of fields, tables, and relationships relevant to the database design schema. For example, table fields could include descriptions as a useful addition to documentation and usage. These are easily viewed within Access Table Design View. See the Preservation section for details on generating the Database Documenter report.

Metadata may be accessed in other ways. Access uses Database Object Listing (MSysObjects) as its underlying standard for its schema. This schema is not used by other platforms, in particular, SQL, which means they are not as interoperable.

- MS Access does NOT use information\_schema standards common to a range of SQL databases, which is one of the standards for reporting information about tables, views, columns, and procedures or calculations in a database. (c.f. [https://en.wikipedia.org/wiki/Information\\_schema](https://en.wikipedia.org/wiki/Information_schema)).
- Access also uses its own SQL standard, [Access SQL](#), which may be incompatible with other SQL standards such as Transaction-SQL.  
<http://rogersaccessblog.blogspot.com/2013/05/what-are-differences-between-access-sql.html>

## Preservation

In this section, we present the following recommendations for basic preservation of MS Access databases. These consist of 1) exporting tables and other objects into interoperable formats, and 2), Making extensive use of the Database Documenter and Relationship Tool to export contributor- and machine-generated metadata. Finally, additional preservation avenues developed by 3<sup>rd</sup> parties for more advanced preservation are presented.

### Exporting Access Tables and Objects

Following the ICPSR recommendations, at minimum, the database tables should be exported into flat CSV files and included alongside the original database itself. To do this, first make a copy of the database and then

1. In Access (2016 used as an example), make sure to group by Object Type and display All Access Objects in the side bar (Fig. 2a).
2. Right click on a table and select Export -> Text File.
3. Enter a name and location for the file. Make sure to change the extension to 'csv' instead of 'txt'. Do NOT check the box to 'Export data with formatting and layout'. Click Ok.
4. In the Export Text Wizard window show in Fig 2b, make sure Delimited is selected.
5. Click the Advanced button in Fig 2b to change date export format and other settings as shown in Fig. 2c. Save the export settings to apply the date export to other tables more quickly.
6. Repeat 2-5 for every table in the database. Also see Linked Tables below for how to deal with linked tables. Note: there is no built-in way to export more than one table at a time.

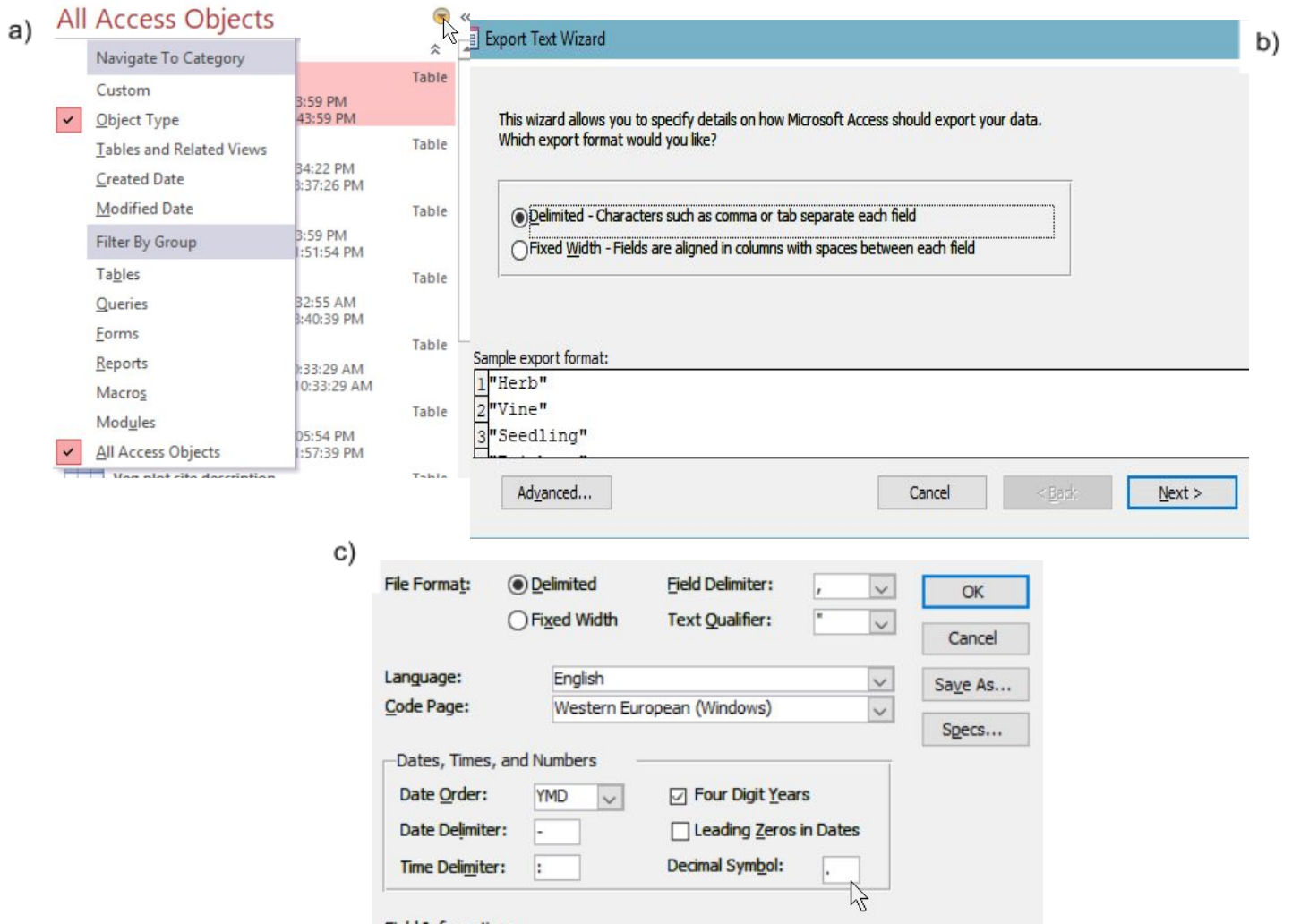


Figure 2: Exporting tables manually into CSV

In addition to tables, other types of objects can also be exported, and they should be when feasible/useful.

- Forms: Can be exported directly to PDF by right-clicking and selecting Export -> PDF or XPS. Bound forms that are directly linked to a table will be exported in their entirety, including the records themselves.
- Macros cannot be exported unless they are first converted to Visual Basic (use the Convert Macros to Visual Basic button in the macro Design tab in Access)
- Modules can be directly exported as text files.
- Queries: Don't use the export menu as this will export the query results instead of the query itself. To export the query, open it in Access in Design View and from the View menu, select SQL View. The text can then be copied into an empty text file with extension '.sql'

## Database Documenter

In addition, it is strongly recommended that the output of the Database Documenter be included alongside the exported tables/objects and any other documentation. The Database Documenter report includes detailed information on each table (data types, any user-entered metadata, etc.) and other Access objects. It also

includes detailed information on the relationships in the database. In the report, each relationship is identified individually using the two table names and the two fields involved, as well as the type of the relationships and any attributes (e.g., whether the relationship is enforced when entering data or whether the deletion of a record cascades to associated tables) as shown in Fig. 3. This information is critical so as to enable understanding the relationships between the exported tables as, upon export, the relationship information is lost.

**Relationships**

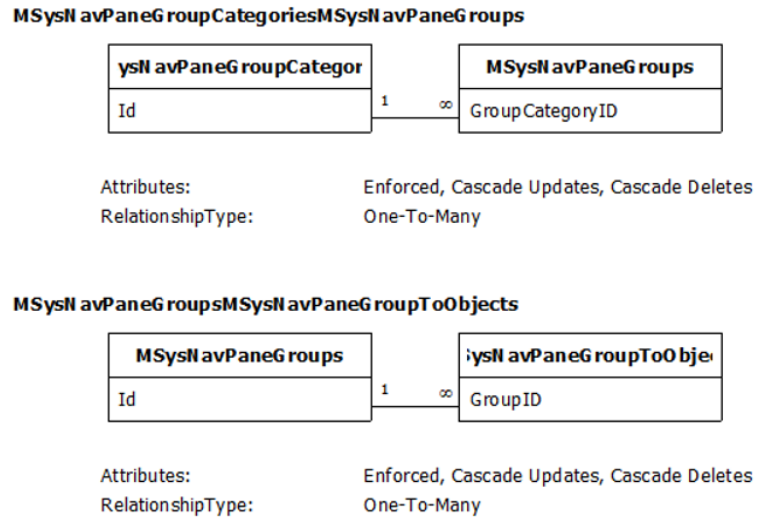


Figure 3: Relationship information generated by Database Documenter

To generate the Database Documenter Report:

1. Click on Database Tools in the ribbon and open the Database Documenter
2. Click the All Object Types tab and select all
3. In the Options, set the settings as shown in Fig. 4 so that all available comments are included.
4. When complete, the report will appear in a Print Preview window. Export the report to PDF by clicking the appropriate button on the ribbon. Note: There are other export options but the PDF export is best for preservation since it is a self-contained document.

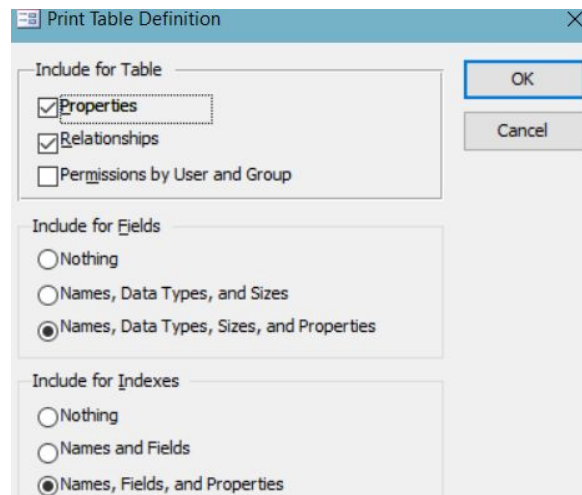


Figure 4: Options for the Database Documenter. The settings shown will capture all available information that is of use to re-users.



Finally, in addition to the Relationships Report generated by the Database Documenter, the diagram from the Relationships editor should be included as a standalone attachment in the form of an image. This is because the relationships report generated by the Database Documenter does not include an overview of all the relationships in the database. To obtain the diagram click on the Relationships button in the Database Tools tab (Fig. 5a) and you will see a window that details the relationships within the database (Fig 5b). You can take a screenshot or export the report to PDF. Typically, the database contributor, rather than the curator, would arrange the relationship diagram in a way that makes sense, especially if intending the shared database to be repurposed.

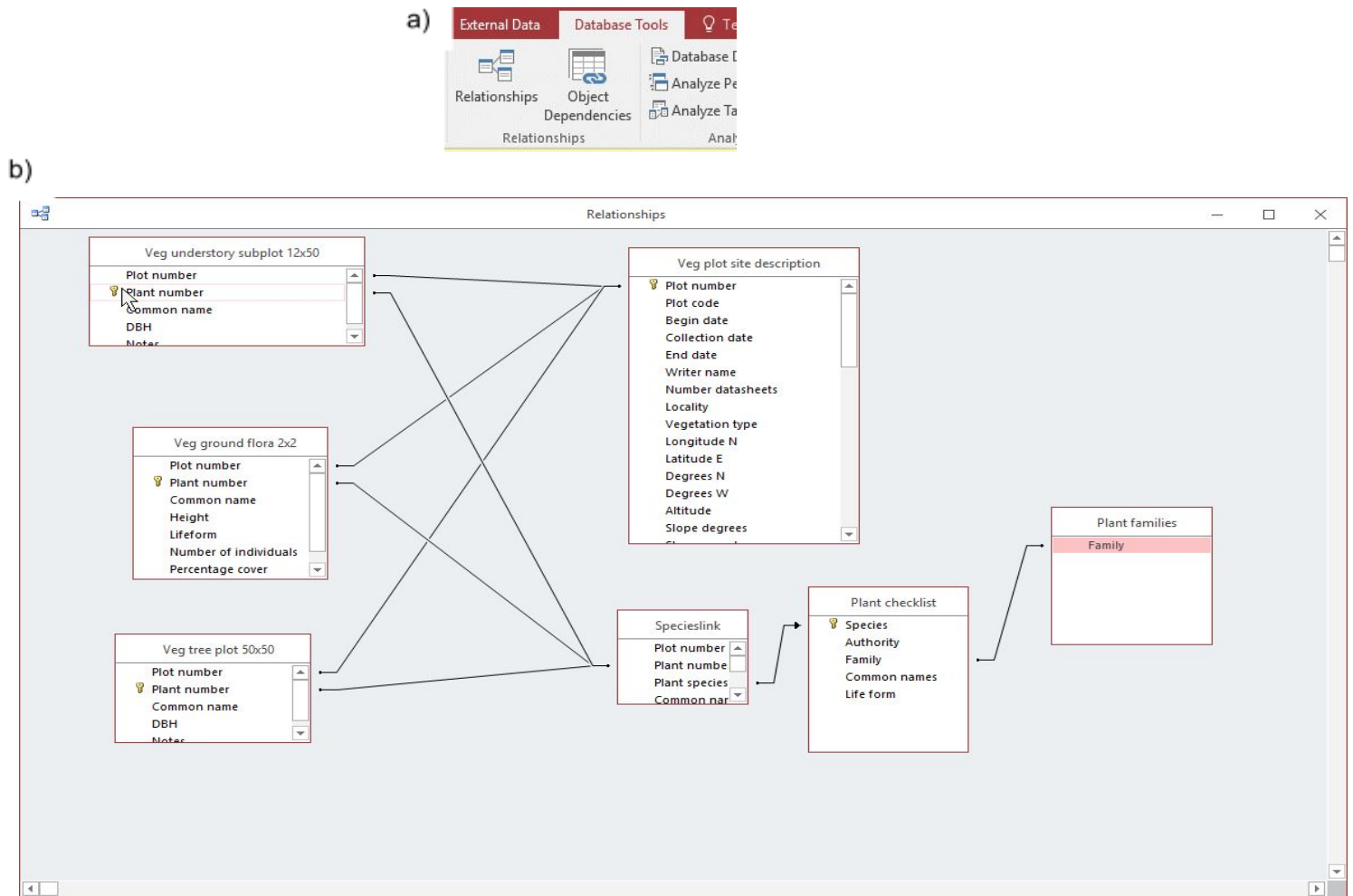


Figure 5: The output of the relationship tool is a useful companion to the Database Documenter output (see Fig. 3)

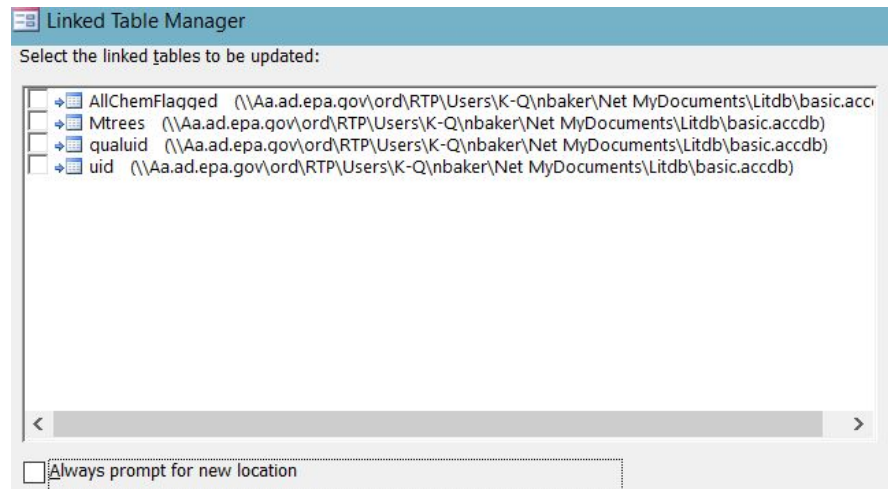


Figure 6: Linked table manager

## Linked Tables

Linked tables are tables that exist in another Access database. If the table is not included with the database being curated, they cannot be opened. These tables can be converted to local tables by right clicking the table and selecting the appropriate menu entry (assuming the linked table was part of the deposit as a separate file or otherwise made available). You can also use the Linked Table Manager (Fig. 6) to see and update all linked tables. Additionally, there is a consideration if a query in Access references external databases (e.g., SQL servers)

## ICPSR Recommendations

ICPSR preservation recommendations: “To preserve relational databases, users should export the database tables as flat rectangular files and preserve the table relationships using, for instance, SQL schema statements. When databases are used as survey instruments or other data input/out mechanisms, the look and feel of the user interface can be preserved by creating a static PDF image of the interface. Promising software is currently under development to normalize relational databases into non-proprietary formats such XML.” (<https://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/chapter4.html>)

## Database Preservation Toolkit (DBPTK) and SIARD standard

The Database Preservation Toolkit facilitates digital preservation of databases by converting a range of platforms to preservation-ready formats, in particular, SIARD 2.0. This format currently supports MS Access 2000 to 2016 (both MDB and ACCDB) formats. It is based on four internationally recognized standards: XML, SQL:2008, UNICODE and ZIP64format. The SAIRD format and toolset have been adopted in Switzerland and other countries. Further information and the executable application can be found at <http://www.database-preservation.com/>. Note: The DBPTK is command-based and would take time to install and implement. The decision to convert to SIARD would be based on your archive’s long-term preservation policy rather than shorter-term reuse for research, for which depositing the original Access format may be sufficient.

## What to look for to make sure this file meets FAIR principles

- Make sure to verify (either by testing each table individually or asking the researcher) that the database does not contain tables that are linked to other databases. In recent version of Access it is indicated by the small arrow beside the table icon (Fig. 7).

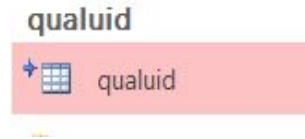


Figure 7: the small arrow to the left of the table indicates it is a linked table

Trying to open these linked tables may result in an error if the curator does not have authenticated access to the remote server or if the linked table no longer exists on the remote server. See the [Curation](#) section for more information.

- Exporting tables to CSV will ensure at least the data is accessible in a non-proprietary format

## Unresolved Issues/Further Questions

- Can LibreOffice (open-source alternative to Microsoft Office) open Access databases directly or can the Access database be exported to a format that can be opened by LibreOffice Base?
  - No, but tables can be linked between Base and Access. [See here for details.](#)
- Preserving databases that reside on servers. How to download (do they act as independent files?) (Sharepoint?) Similarly, Access may be the front-end to tables on server-based SQL, which may be challenging to export and keep associated as a single package for archiving.
- Adding comments to macros and modules (e.g., within the coding).
- Investigating 3rd party tools for accessing Access databases for obtaining version numbers, counts of tables, existence of linked tables without having to manually open each table and query in Access. There are some tools for \*nix like [mdbtools](#) that claim to be able to interact with Access databases.
- Converting Access to MySQL as an alternative open format. Guidelines can be found [at this link](#). Bullzip offers an [Access to MySQL converter utility](#).

## Documentation of curation process: What do capture from curation process

- Exporting tables to CSV: did curator do it or were they provided by the researcher?
- Documenting the presence of linked tables or queries that reference external servers
- Version of MS Access needed to open file
- Any changes to the internal metadata. E.g., if the curator added descriptions to fields or modules using MS Access built-in functionality. This would affect what is produced by the Database Documenter.

- Document relationship among any files or parts exported from the database, e.g., individual tables and the keyfields that relate them, spatial data, reports. Document whether the curator or the researcher generated them.
- Stating whether the curator generated the Database Documenter report vs being generated by the researcher. This implies differences in the quality of the documentation - especially if the curator-generated report was not vetted by the researcher. (A researcher-generated report does not necessarily imply quality either.)

## Bibliography

- Data Carpentry SQL discussion
- Library of Congress has some good descriptions
  - [MDB, ACCDB](#)
- Database preservation: more advanced techniques that are platform-independent
  - [Database Toolkit](#)
  - [SIARD](#) (Software Independent Archiving of Relational Databases)
- [Database Visualization Toolkit](#): Relational Database Viewer for databases based on SIARD 2
- MS Access Documentation and Backup
  - [Create an Access app package](#): for saving as a Snapshot for preservation
  - Database Documenter: [accessing and printing the database design](#)

## Appendix A – MS Access CURATED checklist

### Checklist of CURATED Steps Performed by the DCN

- C** **Check** files and read documentation (risk mitigation, file inventory, appraisal/selection)
- U** **Understand** the data (or try to), if not... (run files/environment, QA/QC issues, readmes)
- R** **Request** missing information or changes (tracking provenance of any changes and why)
- A** **Augment** metadata for findability (DOIs, metadata standards, discoverability)
- T** **Transform** file formats for reuse (data preservation, conversion tools, data viz)
- E** **Evaluate** for FAIRness (licences, responsibility standards, metrics for tracking use)
- D** **Document** your curation activities

### CHECK Step

CURATE Action	Curator Checklist
<p><b>Check</b> data files and read documentation</p> <ul style="list-style-type: none"> <li>• Review the content of the data files (e.g., open and run the files or code).</li> <li>• Verify all metadata provided by the author and review the available documentation.</li> </ul>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Files open as expected                             <ul style="list-style-type: none"> <li><input type="checkbox"/> Issues _____</li> </ul> </li> <li><input type="checkbox"/> Code (modules, macros, queries) runs as expected                             <ul style="list-style-type: none"> <li><input type="checkbox"/> Produces minor errors</li> <li><input type="checkbox"/> Does not run and/or produces many errors</li> <li><input type="checkbox"/> Did not try to run code</li> </ul> </li> <li><input type="checkbox"/> Metadata quality is rich, accurate, and complete                             <ul style="list-style-type: none"> <li><input type="checkbox"/> Metadata has issues _____</li> </ul> </li> <li><input type="checkbox"/> Documentation Type (circle)                              Readme / Codebook / Data Dictionary / Other: _____                             <ul style="list-style-type: none"> <li><input type="checkbox"/> Missing/None</li> <li><input type="checkbox"/> Needs work</li> </ul> </li> </ul>

	<input type="checkbox"/> Human subjects data present <ul style="list-style-type: none"> <li><input type="checkbox"/> Request consent form / participation agreement</li> </ul>
--	--

## UNDERSTAND Step

CURATE Action	Curator Checklist
<p><b>Understand</b> the data (or try to)</p> <ul style="list-style-type: none"> <li>• Check for quality assurance and usability issues such as missing data, ambiguous headings, code execution failures, and data presentation concerns.</li> <li>• Try to detect and extract any “hidden documentation” inherent to the data files that may facilitate reuse.</li> <li>• Determine if the documentation of the data is sufficient for a user with similar qualifications to the author’s to understand and reuse the data. If not, recommend or create additional documentation (e.g., a readme.txt template).</li> </ul>	<p><i>Varies based on file formats and subject domain. For example....</i></p> <p>MS Access</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Is there documentation on tables, relationships, queries, etc?                             <ul style="list-style-type: none"> <li><input type="checkbox"/> Exists in separate file</li> <li><input type="checkbox"/> Included in Access comment fields for tables</li> <li><input type="checkbox"/> Other _____</li> </ul> </li> <li><input type="checkbox"/> Can the data be exported to CSVs easily?                             <ul style="list-style-type: none"> <li><input type="checkbox"/> Yes, no information is lost</li> <li><input type="checkbox"/> No, database contains many queries and relationships</li> <li><input type="checkbox"/> No, Database also has forms/code/macros</li> <li><input type="checkbox"/> Other _____</li> </ul> </li> <li><input type="checkbox"/> Is the database self-contained?                             <ul style="list-style-type: none"> <li><input type="checkbox"/> Yes</li> <li><input type="checkbox"/> No, there are linked tables</li> <li><input type="checkbox"/> No, queries access external data sources.</li> </ul> </li> <li><input type="checkbox"/> Is there or can you obtain documentation or information on how the database was used?</li> <li><input type="checkbox"/> Which tables or queries are the relevant ones used in a publication?</li> </ul>

## REQUEST Step

CURATE Action	Curator Checklist
<p><b>Request</b> missing information or changes</p> <ul style="list-style-type: none"> <li>• Generate a list of questions for the data author to fix any errors or issues.</li> </ul>	<p><i>Narrative describing the concerns, issues, and needed improvements to the data submission.</i></p>

## AUGMENT Step

CURATE Action	Curator Checklist
<p><b>Augment</b> the submission</p> <ul style="list-style-type: none"> <li>• Enhance metadata to best facilitate discoverability.</li> <li>• Create and apply metadata for the data record, including descriptive keywords.</li> <li>• When appropriate, structure and present metadata in domain-specific schemas to facilitate interoperability with other systems.</li> </ul>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Discoverability sufficient             <ul style="list-style-type: none"> <li><input type="checkbox"/> Recommend (circle one) full-text index / file rename / file reorder / file descriptions / zip files into one archive Other _____</li> </ul> </li> <li><input type="checkbox"/> Keywords Sufficient             <ul style="list-style-type: none"> <li><input type="checkbox"/> Suggestions _____</li> </ul> </li> <li><input type="checkbox"/> Linkages Sufficient             <ul style="list-style-type: none"> <li><input type="checkbox"/> Link to report/paper</li> <li><input type="checkbox"/> Link to related data sets</li> <li><input type="checkbox"/> Link to source data</li> <li><input type="checkbox"/> Link to other _____</li> </ul> </li> </ul>

## TRANSFORM Step

CURATE Action	Curator Checklist
<p><b>Transform</b> file formats</p> <ul style="list-style-type: none"> <li>• Identify specialized file formats and their restrictions (e.g., Is the software freely available? Link to it or archive it alongside the data).</li> </ul>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Preferred file formats in use             <ul style="list-style-type: none"> <li><input type="checkbox"/> Recommend conversion from _____ to _____</li> <li><input type="checkbox"/> Retain original formats</li> </ul> </li> <li><input type="checkbox"/> Software needed is readily available             <ul style="list-style-type: none"> <li><input type="checkbox"/> Unclear version of software</li> </ul> </li> </ul>

<ul style="list-style-type: none"> <li>• Transform files into open, non-proprietary file formats that broaden the potential audience for reuse and ensure that preservation actions might be taken by the repository in later steps. Retain original files if data transfer is not perfect.</li> </ul>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Unclear software used</li> <li><input type="checkbox"/> Visualization of data easily accessible             <ul style="list-style-type: none"> <li><input type="checkbox"/> Recommend graphical representation _____</li> <li><input type="checkbox"/> Recommend web-accessible surrogate _____</li> </ul> </li> </ul>
--	--

## EVALUATE Step

CURATE Action	Curator Checklist
<p><b>Evaluate</b> and rate the overall data record for FAIRness.*</p> <ul style="list-style-type: none"> <li>• Score the dataset and recommend ways to increase the FAIRness of the data and become “DCN approved.”</li> </ul>	<p>Findable -</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Metadata exceeds author/ title/ date,</li> <li><input type="checkbox"/> Unique PID (DOI, Handle, PURL, etc.).</li> <li><input type="checkbox"/> Discoverable via web search engines.</li> </ul> <p>Accessible -</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Retrievable via a standard protocol (e.g., HTTP).</li> <li><input type="checkbox"/> Free, open (e.g., download link).</li> </ul> <p>Interoperable -</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Metadata formatted in a standard schema (e.g., Dublin Core).</li> <li><input type="checkbox"/> Metadata provided in machine-readable format (OAI feed).</li> </ul> <p>Reusable -</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Data include sufficient metadata about the data characteristics to reuse</li> <li><input type="checkbox"/> Contact info displayed if the direct assistance of the author needed.</li> <li><input type="checkbox"/> Clear indicators of who created, owns, and stewards the data.</li> <li><input type="checkbox"/> Data are released with clear data usage terms (e.g., a CC License).</li> </ul>

\* Rubric evaluating the FAIR principles are based on the scoring matrix by Dunning, de Smaele, & Böhmer (2017).



# Document

CURATE Action	Curator Checklist
<p><b>Document</b> throughout curation activities.</p> <ul style="list-style-type: none"><li>• Record all necessary information capturing who did what to the dataset and when</li></ul>	<ul style="list-style-type: none"><li><input type="checkbox"/> Accessioning &amp; deposit records (Names, dates, contact information, submission agreements, etc)</li><li><input type="checkbox"/> Repository collection metadata</li><li><input type="checkbox"/> Provenance logs</li><li><input type="checkbox"/> Service workflow</li><li><input type="checkbox"/> Preservation packaging</li><li><input type="checkbox"/> Any additional requirements at your institution</li></ul>