Understanding First Normal Form

First normal form is the first rule to apply. It can be demonstrated by using an Excel spreadsheet that contains information about authors, books, and publishers (see Figure 9.29). In the spreadsheet, Dan Brown is listed as an Author in columns C and D; the titles of the books that Brown authored are listed on row 5 in column D. The remaining fields contain additional information related to the books. This example would not pass the 1NF test because column E—including the books that Brown authored—has multiple values in a single cell. Brown, the author, is listed once, and the books he wrote are listed in one cell. Even titles on two lines in the same cell are considered multiple values in one cell. If this Excel spreadsheet were imported into Access, the corresponding Access table with repeating groups would be as shown in Figure 9.30. This table shows the same condition with multiple values in the Title field. Repeating groups are not allowed in a normalized table.

	A	B	C	D	E	F	G
1	ISBN	AuthorID	AuthorFirst	AuthorLast	Title	Publisher	PubYear
2	9781401208417	ALMO01	Alan	Moore	V for Vendetta	DC Comics	2005
3	9780767931557	BEME01	Ben	Mezrich	The Accidental Billionaires	Anchor Books	2009
4	0805029648 / 9780399154379	DABA01	Dave	Barry	Bad Habits / History of the Millenium (So Far)	Henry Holt & Co. / G. P. Putman's Sons	1987 / 2007
5	0312995423 / 9780593054277	DABR01	Dan	Brown	Digital Fortress / The Lost Symbol	St. Martin / Doubleday	1998 / 2009
6	0809400774	DABR02	Dale	Brown	American Cooking: The Northwest	Time Life	1970
7	9780307405807	DOSA01	Douglas	Sarine	The Ninja Handbook	Three Rivers Press	2008
8	9781572439597	JAST01	Jayson	Stark	The Stark Truth	Triumph Books	2007
9	0380788624	NEST01	Neal	Stephenson	Cryptonomicon	Perennial	1999
10	9780345517951	ROJA01	Ron	Jaworski	The Games that Changed the Game	ESPN Books	2010

(Two values in one cell — annotation pointing to row 5, columns D and E)

Access 2016, Windows 10, Microsoft Corporation

FIGURE 9.29 Poorly Designed Excel Spreadsheet

9780767931557	BEME01	Ben	Mezrich	The Accidental Billionaires	Anchor Books	2009
080502964897 / 80399154379	DABA01	Dave	Barry	Bad Habits / History of the Millenium (So Far)	Henry Holt & Co. / G. P. Putman's Sons	1987 / 2007
031299542397 / 80593054277	DABR01	Dan	Brown	Digital Fortress / The Lost Symbol	St. Martin / Doubleday	1998 / 2009
0809400774	DABR02	Dale	Brown	American Cooking: The Northwest	Time Life	1970
9780307405807	DOSA01	Douglas	Sarine	The Ninja Handbook	Three Rivers Press	2008
9781572439597	JAST01	Jayson	Stark	The Stark Truth	Triumph Books	2007
0380788624	NEST01	Neal	Stephenson	Cryptonomicon	Perennial	1999
9780345517951	ROJA01	Ron	Jaworski	The Games that Changed the Game	ESPN Books	2010

(Two values in one record — annotation pointing to the DABR01 row)

Access 2016, Windows 10, Microsoft Corporation

FIGURE 9.30 Poorly Designed Access Database

Implement First Normal Form

STEP 1 »» A table is defined as being in ***first normal form (1NF)*** if it contains no repeating groups or repeating columns. In the book example, the Title column might be replaced with Title1, Title2, and Title3. This might appear to correct the repeating groups problem because each cell contains one piece of information; however, this type of design still violates 1NF and must be corrected. In this example, it is still difficult to find data efficiently.

Violations of the 1NF rule come in many forms. For example, you cannot have multiple values in the same field. Each of the following would be violations of 1NF:

- An Authors field listing multiple authors separated by commas like *Poatsy, Cameron, Williams*

- An Address2 field listing the city, state, and ZIP together such as *Paterson, NJ 07501*

- An Instructor field that contains the following values for the same class on two lines:

  Poatsy

  Cameron

Likewise, you cannot have multiple columns in the same record that store the same sort of data. A table with an Author1 field set to *Poatsy*, Author2 field set to *Cameron*, and Author3 field set to *Williams* is a violation of 1NF.

Each of those examples has a common thread: They make it difficult to find specific pieces of information, and that is the essence of the repeating groups problem. Tables containing repeating groups make it more difficult to add new entries, update existing entries, and properly extract information when running queries. For example, to see the information for *Digital Fortress*, you could create a filter or query searching for those results. However, Access considers them to be part of the same record, and you would also see the book *The Lost Symbol*. As a matter of fact, when you attempt to filter your data (as shown in Figure 9.31), two book names are listed as one piece of data. Similar problems would exist with queries searching for a specific book. In other words, there is no easy way to only show the information for *Digital Fortress*. This occurs because all the books by one author are contained in the same record.
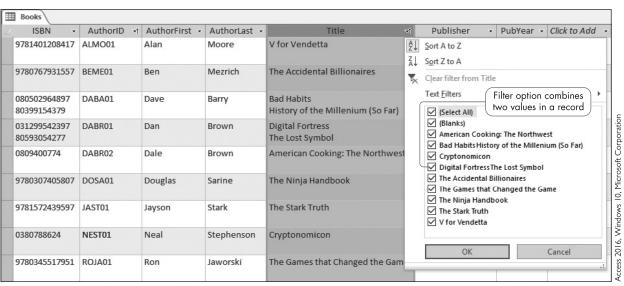

FIGURE 9.31 Filter Issues with Repeating Groups

To fix a table that has repeating groups, change the table so each unique piece of information is stored on a separate record in the same field. Any columns should have the missing information added, even if this leads to repetition. You can do this in Excel before importing to Access, or alternately add rows in Access and separate the repeating data into individual rows.

Figure 9.32 shows the Books table in Access after it has been put into first normal form; the single Brown row has been split into two rows, and the books by the same author have been separated. Now each book has its own record.

Each book has a record

ISBN	AuthorID	AuthorFirst	AuthorLast	Title	Publisher	PubYear
9781401208417	ALMO01	Alan	Moore	V for Vendetta	DC Comics	2005
9780767931557	BEME01	Ben	Mezrich	The Accidental Billionaires	Anchor Books	2009
0805029648	DABA01	Dave	Barry	Bad Habits	Henry Holt & Co.	1987
9780399154379	DABA01	Dave	Barry	History of the Millenium (So Far)	G. P. Putman's Sons	2007
0312995423	DABR01	Dan	Brown	Digital Fortress	St. Martin	1998
9780593054277	DABR01	Dan	Brown	The Lost Symbol	Doubleday	2009
0809400774	DABR02	Dale	Brown	American Cooking: The Northwest	Time Life	1970
9780307405807	DOSA01	Douglas	Sarine	The Ninja Handbook	Three Rivers Press	2008
9781572439597	JAST01	Jayson	Stark	The Stark Truth	Triumph Books	2007
0380788624	NEST01	Neal	Stephenson	Cryptonomicon	Perennial	1999
9780345517951	ROJA01	Ron	Jaworski	The Games that Changed the Game	ESPN Books	2010

FIGURE 9.32 Books Table in First Normal Form (1NF)

At this point, although the table is in 1NF, it still has redundancy issues. In fact, redundancy has been introduced into the table. However, because normalization is a process, further normalization will remove the introduced redundancies.

There are many other ways data can violate 1NF. Note that even database professionals may struggle with normalization, so your goal should be to become familiar with the process.

Understanding Second Normal Form

Most bookstore databases contain information about location and quantity in stock. Therefore, the earlier example has been expanded to include three extra fields, as shown in Figure 9.33. For the purposes of this example, assume the bookstore has two locations, one in the town of Paterson and another one in the town of Wanaque. Because this more realistic version of the table has the location added, a problem now exists. Notice in Figure 9.33 that the ISBN cannot be the primary key for this table because the same book (with the same ISBN) can appear in both locations. In this case, a good option is to use a combination of two fields as the primary key. The composite key for this table would be ISBN plus the location, which uniquely identifies each row.

ISBN	AuthorID	AuthorFirst	AuthorLast	Title	Publisher	PubYear	Location	OnHand	Aisle
9781401208417	ALMO01	Alan	Moore	V for Vendetta	DC Comics	2005	Paterson	2	5
9780767931557	BEME01	Ben	Mezrich	The Accidental Billionaires	Anchor Books	2009	Paterson	11	5
9780767931557	BEME01	Ben	Mezrich	The Accidental Billionaires	Anchor Books	2009	Wanaque	6	5
0805029648	DABA01	Dave	Barry	Bad Habits	Henry Holt & Co.	1987	Wanaque	1	4
9780399154379	DABA01	Dave	Barry	History of the Millenium (So Far)	G. P. Putman's Sons	2007	Wanaque	1	2
0312995423	DABR01	Dan	Brown	Digital Fortress	St. Martin	1998	Wanaque	3	1
0312995423	DABR01	Dan	Brown	Digital Fortress	St. Martin	1998	Paterson	2	5
9780593054277	DABR01	Dan	Brown	The Lost Symbol	Doubleday	2009	Paterson	11	3
9780593054277	DABR01	Dan	Brown	The Lost Symbol	Doubleday	2009	Wanaque	5	4
0809400774	DABR02	Dale	Brown	American Cooking: The Northwest	Time Life	1970	Paterson	1	8
9780307405807	DOSA01	Douglas	Sarine	The Ninja Handbook	Three Rivers Press	2008	Paterson	1	3
9781572439597	JAST01	Jayson	Stark	The Stark Truth	Triumph Books	2007	Wanaque	1	6
0380788624	NEST01	Neal	Stephenson	Cryptonomicon	Perennial	1999	Paterson	1	1
9780345517951	ROJA01	Ron	Jaworski	The Games that Changed the Game	ESPN Books	2010	Paterson	2	1

Repeated information for a book found in two locations

FIGURE 9.33 Expanded Books Table