



DATABASE SYSTEMS

Transform and upload data



BUDAPESTI MŰSZAKI
ÉS GAZDASÁGTUDOMÁNYI EGYETEM
Építőmérnöki Kar - építőmérnöki képzés 1782 óta

Fotogrammetria és Térinformatika Tanszék

Bence Molnár

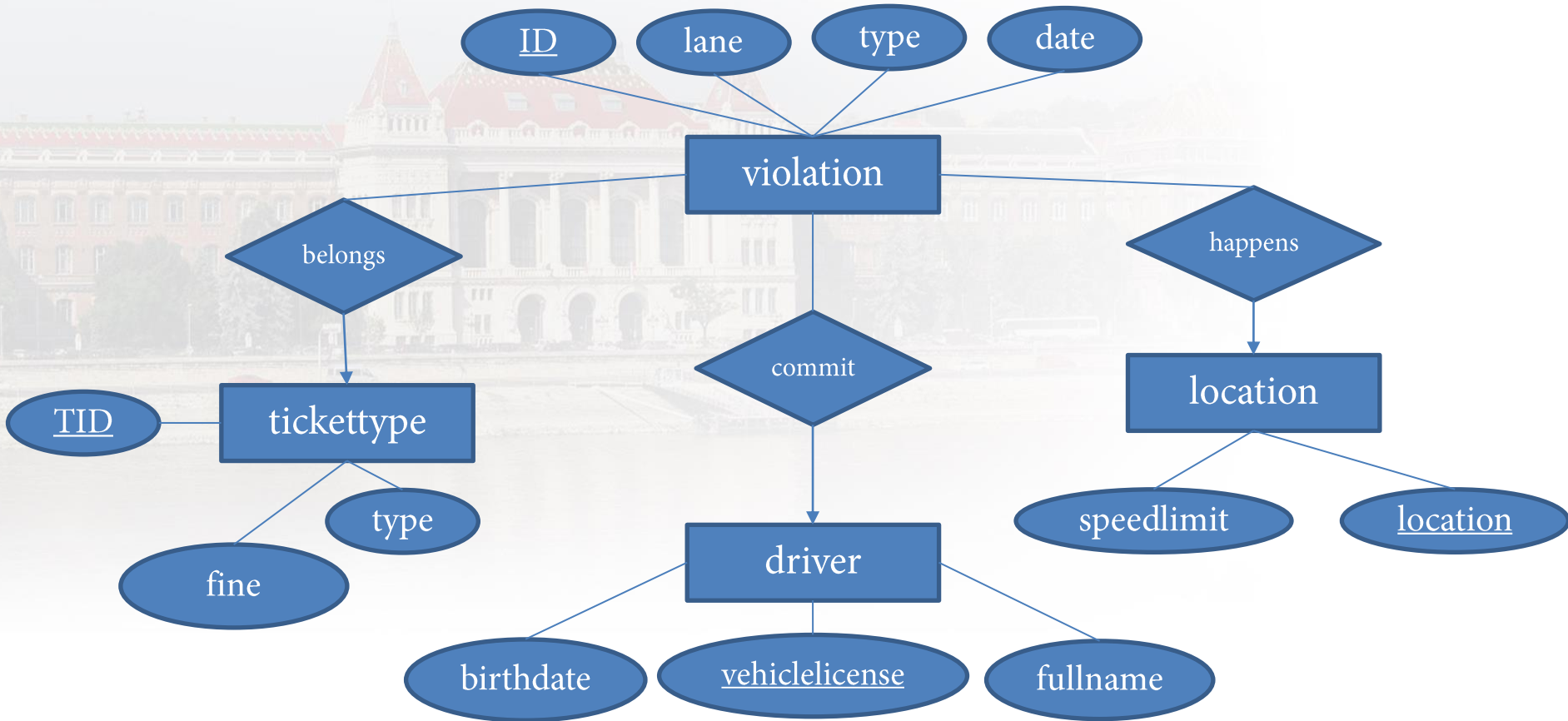
2022.03.28.

AGENDA

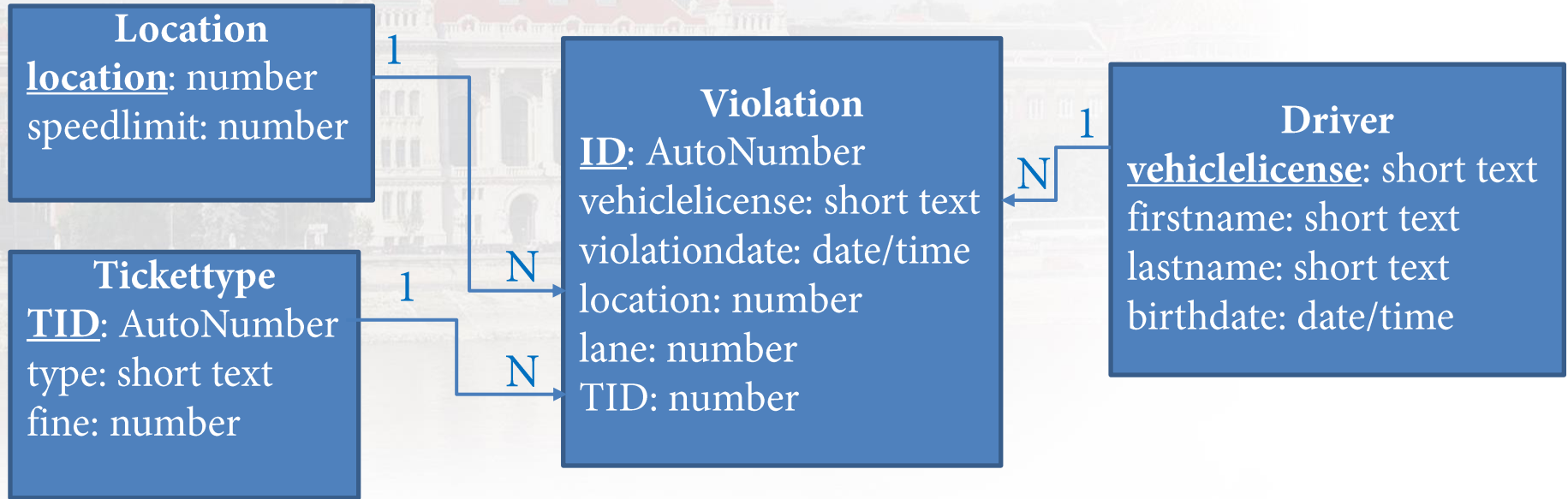
Example: Catalogue of traffic rule violations

- Defining table
- Defining relationships of tables
- Cleaning and transforming data
- Upload data and remove duplicates

CATALOGUE OF TRAFFIC RULE VIOLATIONS – ER DIAGRAM



CATALOGUE OF TRAFFIC RULE VIOLATIONS – RELATIONAL SCHEMA DIAGRAM



CATALOGUE OF TRAFFIC RULE VIOLATIONS

driver	
Field Name	Data Type
vehiclelicense	Short Text
firstname	Short Text
lastname	Short Text
birthdate	Date/Time

General	
Lookup	
Field Size	15
Format	@
Input Mask	LLL\ -000;0;_
Caption	
Default Value	
Validation Rule	
Validation Text	
Required	Yes
Allow Zero Length	No
Indexed	Yes (No Duplicates)
Unicode Compression	No
IME Mode	No Control
IME Sentence Mode	None
Text Align	General

•attribute types of driver table

primary key:

vehiclelicense: short text

field size: 15

input mask: LLL\ -000;0;_

additional fields:

firstname: short text

lastname: short text

birthdate: date/time

input mask: 0000.00.00;0;_

CATALOGUE OF TRAFFIC RULE VIOLATIONS

location	
Field Name	Data Type
location	Number
speedlimit	Number

General	Lookup
Field Size	Long Integer
Format	
Decimal Places	0
Input Mask	
Caption	
Default Value	0
Validation Rule	
Validation Text	
Required	Yes
Indexed	Yes (No Duplicates)
Text Align	General

•attribute types of location table
primary key:

location: number

additional fields:

speedlimit: number

CATALOGUE OF TRAFFIC RULE VIOLATIONS

tickettype	
Field Name	Data Type
TID	AutoNumber
type	Short Text
fine	Number

- attribute types of tickettype table
primary key:

TID: AutoNumber

General	
Field Size	200
Format	
Input Mask	
Caption	
Default Value	
Validation Rule	
Validation Text	
Required	Yes
Allow Zero Length	No
Indexed	No
Unicode Compression	Yes
IME Mode	No Control
IME Sentence Mode	None
Text Align	General

additional fields:

type: short text

field size: 200 characters

fine: number

duplicates are allowed

CATALOGUE OF TRAFFIC RULE VIOLATIONS

tickettype			
	TID	type	fine
+	1	overspeed 0-20 km/h	15000
+	2	overspeed 21-35 km/h	30000
+	3	overspeed 36-50 km/h	45000
+	4	overspeed 51-65 km/h	60000
+	5	overspeed 66-80 km/h	90000
+	6	overspeed 81-95 km/h	130000
+	7	speed without lamp	20000
+	8	unauthorized road use	15000
+	9	expired MOT 3-6 month	15000
*	(New)		

•data of tickettype

CATALOGUE OF TRAFFIC RULE VIOLATIONS

violation	
Field Name	Data Type
ID	AutoNumber
vehiclelicense	Short Text
violationdate	Date/Time
location	Number
lane	Number
TID	Number

- attribute types of violation table
primary key:

ID: AutoNumber

additional fields:

vehiclelicense: short text FK: [driver]

violationdate: date/time

location: number FK:[location]

lane: number

TID: number FK: [tickettype]

General	
Display Control	Combo Box
Row Source Type	Table/Query
Row Source	SELECT tickettype.[TID], [tickettype].type FROM tickettype ORDER BY [TID];
Bound Column	1
Column Count	2
Column Heads	No
Column Widths	0cm;3,556cm
List Rows	16
List Width	3,556cm
Limit To List	Yes
Allow Multiple Values	No
Allow Value List Edits	No
List Items Edit Form	
Show Only Row Source V	No

OPTIONS TO IMPORT

A: Excel source, cleaning and transforming data in Excel, import into a new Access table

- access table: driver

B: Excel source, cleaning and transforming data in Excel, import into existing Access table

C: CSV source and import into Access

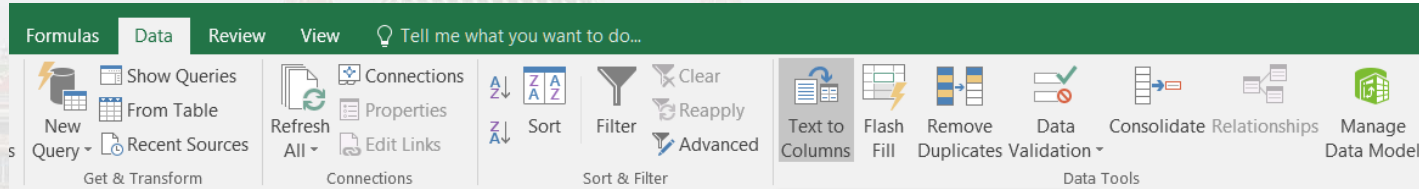
D: Excel source, import, transforming and filtering data in Access

DATA CLEANING AND TRANSFORMING - EXCEL

- acquired data does not fit to designed schema, we need to transform it
 - Date is stored in 6 columns (year, month, day, hour, min, sec)
 - Platenumber is split and stored in two columns
 - Firstnames and lastnames are stored in the same field
 - Type of violation is textual and we need codes instead

DATA CLEANING AND TRANSFORMING - EXCEL

- prepare data for driver table – split names

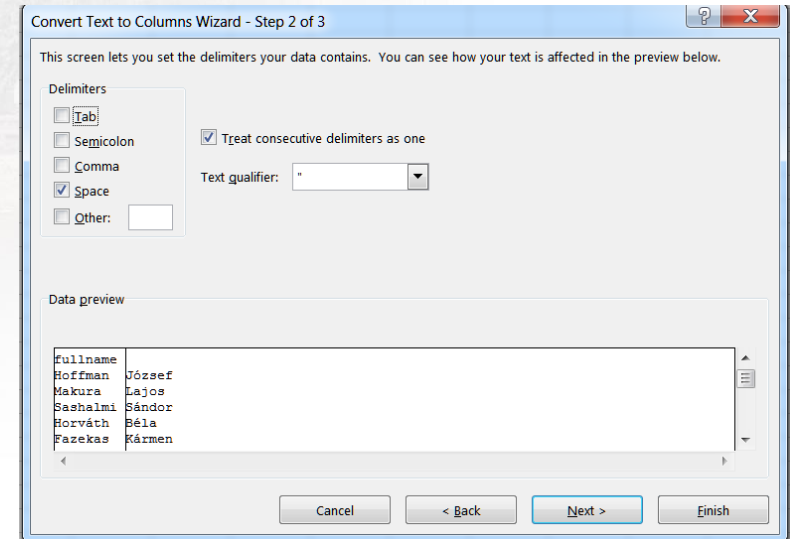


Data tab – text to columns

1. Defining field separator

2. Defining attribute types and order for each column

Potential error: columns are not shifted, but content is overwritten by new items



DATA CLEANING AND TRANSFORMING - EXCEL

- prepare data for driver table – split driver's name

Built-in functions:

FIND, LEN, LEFT, RIGHT

lastname: LEFT(field_of_fullname; FIND(" "; field_of_fullname))

firstname: RIGHT(field_of_fullname; LEN(field_of_fullname)-FIND(" "; field_of_fullname))

Argument separator may differ different machines
(based on locales settings). Options are: , ;

Hint: decimal separator character differs in some
languages

DATA CLEANING AND TRANSFORMING - EXCEL

- prepare data for driver table – join platenumber parts

Built-in functions:

CONCATENATE or & operator

vehiclelicense: CONCATENATE(text1; "-"; text2) text1&"-"&text2

PREPARING DATE

LEFT, RIGHT, CONCATENATE, TEXT functions

Text to date conversion:

Function: VALUE

violationdate: VALUE("2018/3/2 17:35:20")

"YYYY/MM/DD hh:mm:ss"

Excel display format: Custom: m/d/yyyy h:mm:ss

DATA CLEANING AND TRANSFORMING - EXCEL

- prepare data for violation table – search for TID values

Built-in functions:

IF, VLOOKUP

Lookup for TID values for violation

If overspeed applies, we use lower and upper boundaries
and define speed differences.

IF(field_value="overspeed"; speed-speedlimit; 0)

		tickettype
	TID	type
0	20	1 overspeed 0-20 km/h
21	35	2 overspeed 21-35 km/h
36	50	3 overspeed 36-50 km/h
51	65	4 overspeed 51-65 km/h
66	80	5 overspeed 66-80 km/h
81	95	6 overspeed 81-95 km/h
		7 speed without lamp
		8 unauthorized road use
		9 expired MOT 3-6 month

DATA CLEANING AND TRANSFORMING - EXCEL

- prepare data for violation table – TID data

Usage of VLOOKUP: search for values vertically in a given table's first column

VLOOKUP(search_value; lookuprange; column_number; TRUE/FALSE)

TRUE- approximate match

FALSE- exact match

speed without lamp
overspeed

VLOOKUP(\$D2;\$L\$3:\$N\$8;3;TRUE)

39

TID		typ
0	20	1 ove
21	35	2 ove
36	50	3 ove
51	65	4 ove
66	80	5 ove
81	95	6 ove

Hint: approximate match
works only if lookup range is
sorted

DATA CLEANING AND TRANSFORMING - EXCEL

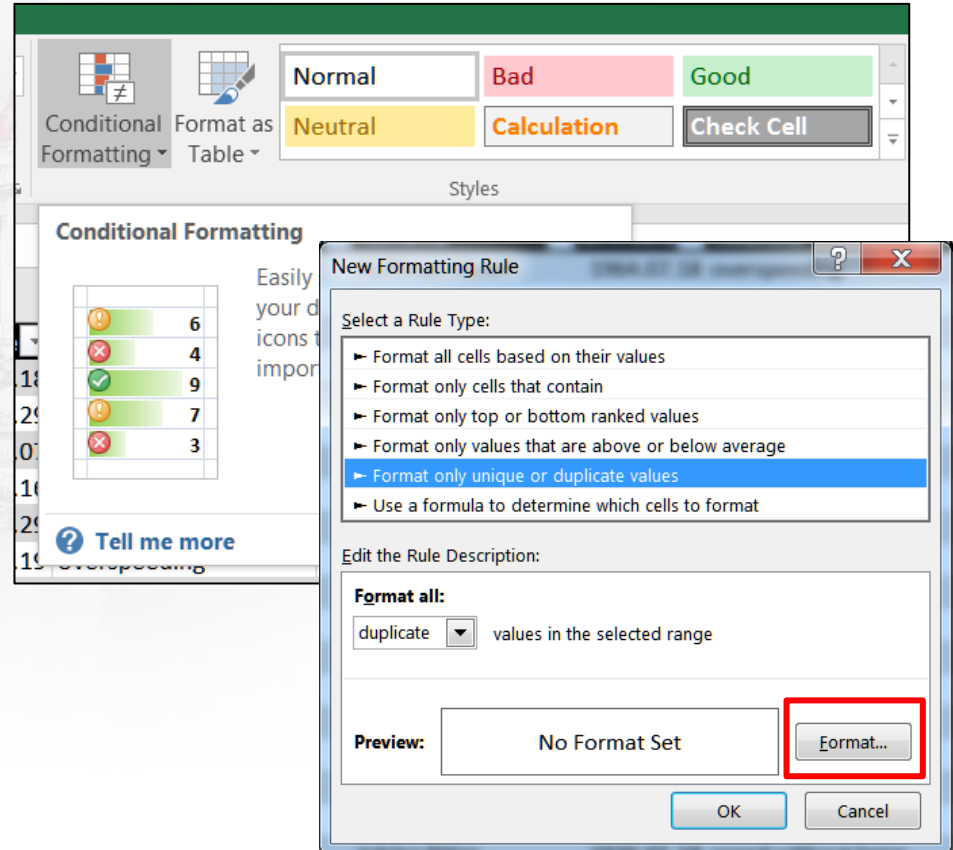
- Search for duplicates

Conditional formatting

Home tab – styles ->

Highlight cell rules ->

Duplicate values

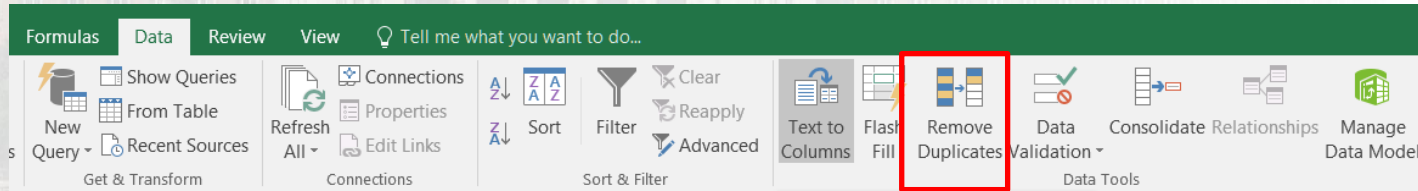


DATA CLEANING AND TRANSFORMING - EXCEL

- remove duplicates

Data tab – data tools

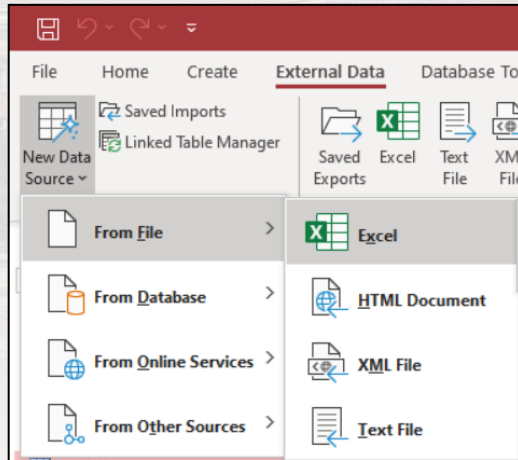
Remove Duplicates



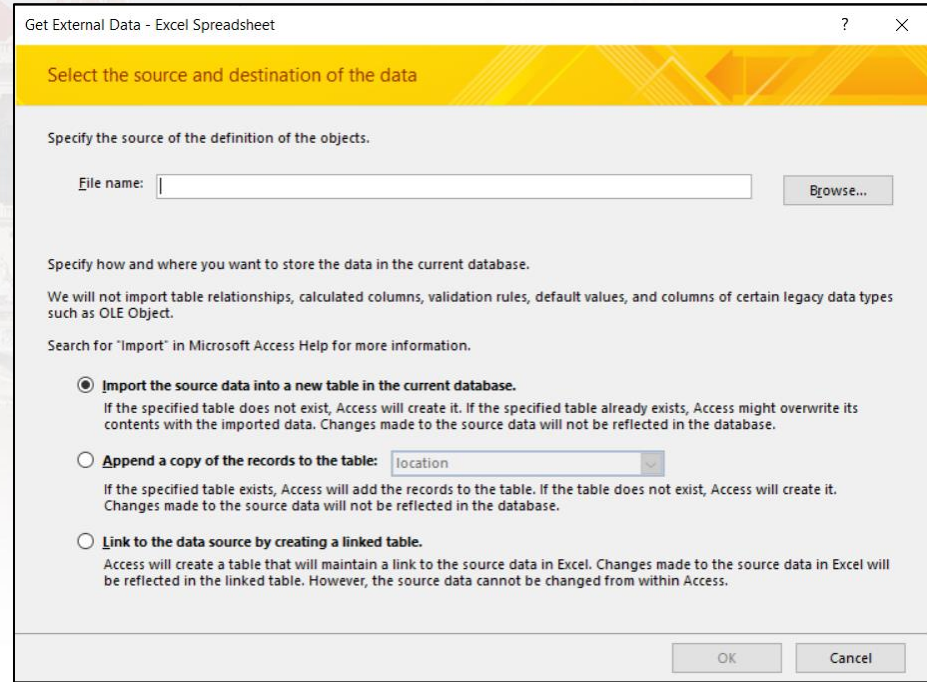
UPLOAD DATA - ACCESS

- import from excel

External Data- import excel



- into a new table
- into existing table (append)
- Link data source (realtime connection)



UPLOAD DATA – ACCESS • NEW TABLE

1. Select worksheet

☒ Show Worksheets
☐ Show Named Ranges

source
driver
prepare_violation
violation
practice_data
error sheet

2. Does source data include headers?

Microsoft Access can use your column headings as field names for your table. Does the first row specified contain column headings?

☐ First Row Contains Column Headings

3. Finetune datatypes for columns

Field Options

Field Name: Data Type:

Indexed: ☐ Do not import field (Skip)

fullname	birthdate	Co
Hoffman József	23576	ov
Makura Lajos	30070	ov
Sashalmi Sándor	23414	ov
Horváth Béla	29875	sp
Fazekas Kármén	24682	ov

4. Define primary key

- ☒ Let Access add primary key.
☐ Choose my own primary key.
☐ No primary key.

OPTIONS TO IMPORT

A: Excel source, cleaning and transforming data in Excel, import into a new Access table

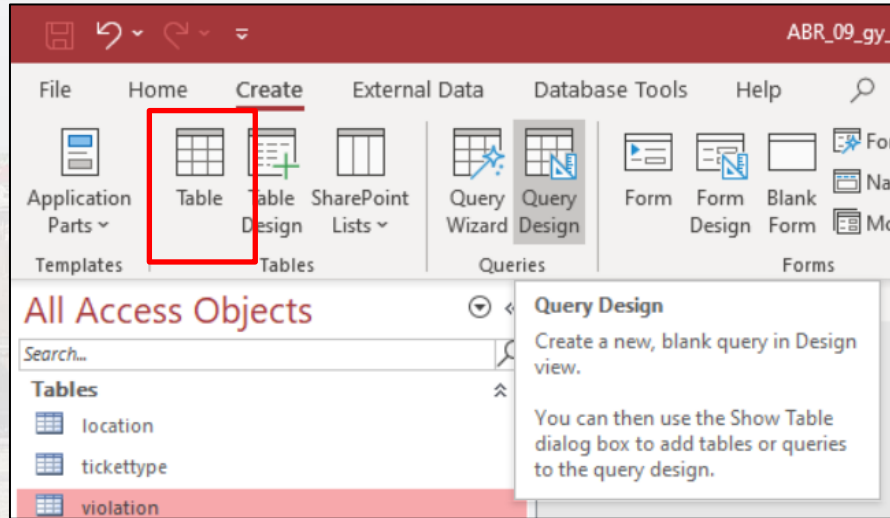
B: Excel source, cleaning and transforming data in Excel, import into existing Access table

- access table: violation

C: CSV source and import into Access

D: Excel source, import, transforming and filtering data in Access

CREATE NEW ACCESS TABLE



- Create new table

UPLOAD DATA – ACCESS • EXISTING TABLE

Get External Data - Excel Spreadsheet

Select the source and destination of the data

Specify the source of the definition of the objects.

File name:

Specify how and where you want to store the data in the current database.

We will not import table relationships, calculated columns, validation rules, default values, and columns of certain legacy data types such as OLE Object.

Search for "Import" in Microsoft Access Help for more information.

☐ **Import the source data into a new table in the current database.**
If the specified table does not exist, Access will create it. If the specified table already exists, Access might overwrite its contents with the imported data. Changes made to the source data will not be reflected in the database.

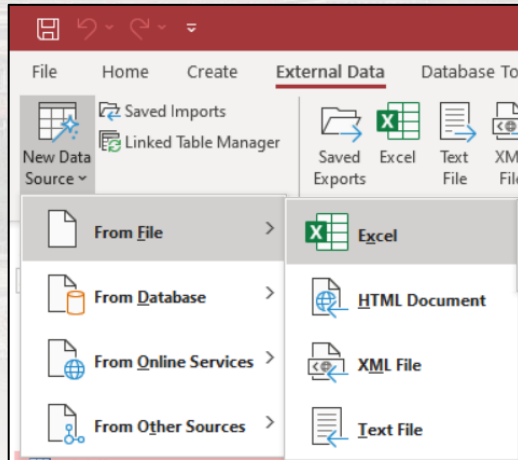
☒ **Append a copy of the records to the table:**
If the specified table exists, Access will add the records to the table. If the table does not exist, Access will create it. Changes made to the source data will not be reflected in the database.

☐ **Link to the data source by creating a linked table.**
Access will create a table that will maintain a link to the source data in Excel. Changes made to the source data in Excel will be reflected in the linked table. However, the source data cannot be changed from within Access.

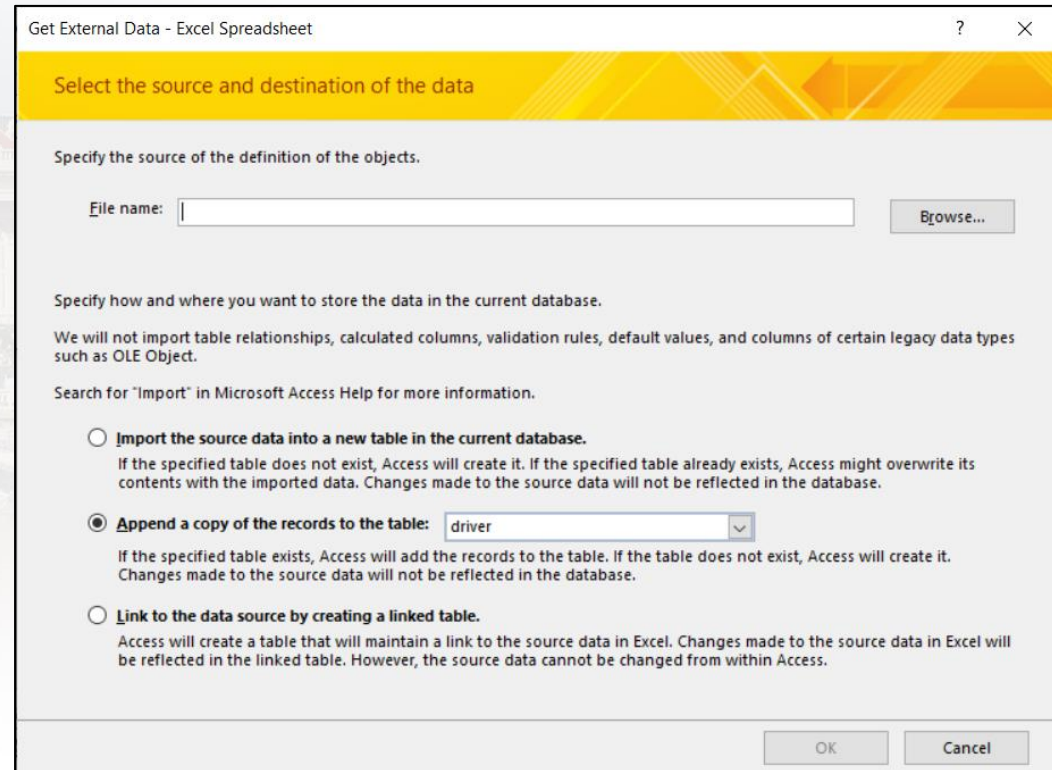
UPLOAD DATA – ACCESS • EXISTING TABLE

- import from Excel

External Data- import Excel



- into existing table



UPLOAD DATA – ACCESS • EXISTING TABLE

- when importing into an existing table, pay attention to the number of columns. If the data you want to read contains more columns than specified in the table, the import will not complete.

- Often the problem is that when you save Excel, the cursor is in an empty field outside the range, and then Access expands the range to be imported when you import it.

Import Spreadsheet Wizard

Microsoft Access can use your column headings as field names for your table. Does the first row specified contain column headings?

☒ First Row Contains Column Headings

	vehiclelicense	firstname	lastname	birthdate
1	GIM-545	József	Hoffman	23576
2	HIS-559	Lajos	Makura	30070
3	HAA-926	Sándor	Sashalmi	23414
4	ILE-402	Béla	Horváth	29875
5	KWC-470	Kármén	Fazekas	24682
6	KPX-527	Zsolt	Csóke	31916
7	LHT-362	Péter	Juhász	27777
8	GAP-636	Zoltán	Kiss	33029
9	FFJ-799	Péter	Bugyik	29894
10	HFD-576	Zsolt	Laufer	33572
11	DJR-964	Márton	Kecskés	27972
12	JZF-138	Klára	Ulmi	24950
13	JHE-498	Dénes	Szabó	29949
14	FEJ-888	Alajos	Keresztes	27141

< [] >

Cancel < Back **Next >** Finish

OPTIONS TO IMPORT

A: Excel source, cleaning and transforming data in Excel, import into a new Access table

B: Excel source, cleaning and transforming data in Excel, import into existing Access table

C: CSV source and import into Access

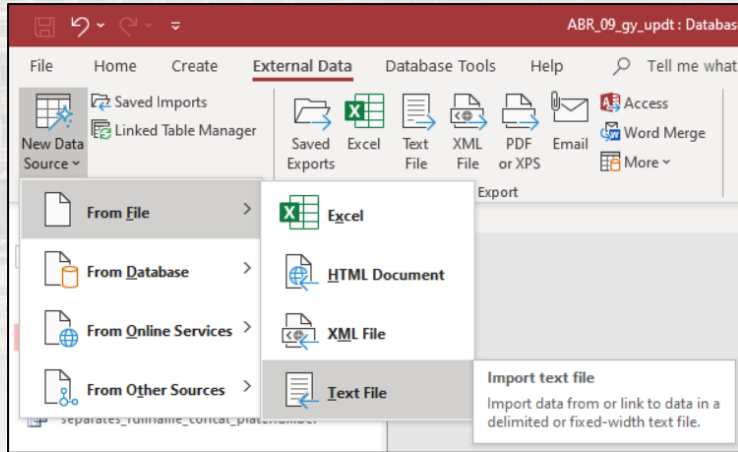
- access table: driver

D: Excel source, import, transforming and filtering data in Access

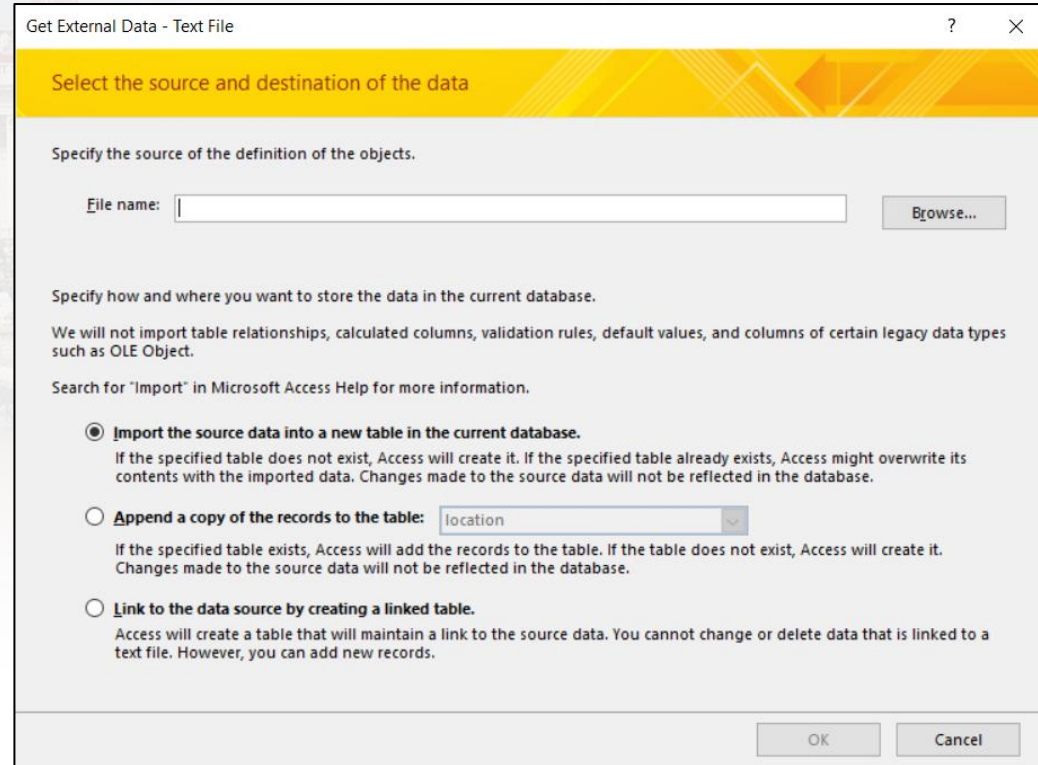
UPLOAD DATA – ACCESS

Import from text file (*.txt, *.csv)

External Data- import text file



- into a new table
- into existing table (append)
- Link data source (realtime)



UPLOAD DATA – ACCESS

1. Code page setting
decimal symbol
language, date and time

ABR_extra_driver_data Import Specification

File Format: ☒ Delimited ☐ Fixed Width

Field Delimiter: {space} Text Qualifier: {none}

Language: Hungarian Code Page: Central European (DOS)

Dates, Times, and Numbers

Date Order: YMD Date Delimiter: / Time Delimiter: : Four Digit Years: ☒ Leading Zeros in Dates: ☐ Decimal Symbol: ,

Field Information:

Field Name	Skip
vehidlicense	☐
firstname	☐
lastname	☐
birthdate	☐
*	☒

Import Text Wizard

Your data seems to be in a 'Fixed Width' format. If it isn't, choose the format that more correctly describes your data.

☒ Delimited - Characters such as comma or tab separate each field

Import Text Wizard

What delimiter separates your fields? Select the appropriate delimiter and see how your text is affected in the preview below.

Choose the delimiter that separates your fields:

☐ Tab ☐ Semicolon ☐ Comma ☒ Space ☐ Other:

☐ First Row Contains Field Names Text Qualifier: {none}

1	2	3	4
ZZN-214	Pál	Krisztián	1982-10-31
RRH-493	Balázs	Vince	1945-02-11
QQM-241	Kocsis	Endre	1941-05-20
ZZE-527	Major	Albert	1956-07-07
JJD-306	Miksa	Márkó	1990-01-08
DDZ-995	Mezei	Gábor	1975-08-05
JJS-428	Egyed	Lajos	1993-01-20
IIT-361	Jakab	Lajos	1933-03-26
OOL-830	Bodnár	Máté	1931-10-12
HHG-250	Szekeres	Bendegúz	1998-11-09
LLA-864	Fehér	Ildikó	1944-10-14
BBQ-727	Szabó	Krisztina	1992-03-28
EEP-681	Kovács	Georgina	1939-08-24
NNC-477	Bács	Júlia	1935-02-05

Advanced...

Central European(DOS)



UPLOAD DATA – ACCESS

2. Field delimiter header information

Import Text Wizard

What delimiter separates your fields? Select the appropriate delimiter and see how your text is affected in the preview below.

Choose the delimiter that separates your fields:

☐ Tab ☐ Semicolon ☐ Comma ☒ Space ☐ Other:

☐ First Row Contains Field Names

Text Qualifier:

ZZN-214	Púl	Krisztián	1982-10-31
RRH-493	Balázs	Vince	1945-02-11
QQM-241	Kocsis	Endre	1941-05-20
ZZE-527	Major	Albert	1956-07-07
JJD-306	Miksa	Márkó	1990-01-08
DDZ-995	Mezei	Gábor	1975-08-05
JJS-428	Egyed	Lajos	1993-01-20
IIT-361	Jakab	Lajos	1933-03-26
OOL-830	Bodnár	Máté	1931-10-12
HHG-250	Szekeres	Bendegúz	1998-11-09
LLA-864	Fehér	Ildikó	1944-10-14
BBQ-727	Szabó	Krisztina	1992-03-28
EEP-681	Kovács	Georgina	1939-08-24
NNC-477	Bács	Júlia	1935-02-05

<

UPLOAD DATA – ACCESS

3. Finetune data type – only available upon importing into new table

Import Text Wizard

You can specify information about each of the fields you are importing. Select fields in the area below. You can then modify field information in the 'Field Options' area.

Field Options

Field Name: Data Type:

Indexed: ☐ Do not import field (Skip)

plate	name1	name2	birthdate
ZZN-214	Pál	Krisztián	1982-10-31
RRH-493	Balázs	Vince	1945-02-11
QQM-241	Kocsis	Endre	1941-05-20
ZZE-527	Major	Albert	1956-07-07
VJD-306	Miksa	Márkó	1990-01-08
DDZ-995	Mezei	Gábor	1975-08-05
VJS-428	Egyed	Lajos	1993-01-20
IIT-361	Vakab	Lajos	1933-03-26
OOL-830	Bodnár	Máté	1931-10-12
HHG-250	Szekeres	Bendegúz	1998-11-09
LLA-864	Fehér	Ildikó	1944-10-14
BBQ-727	Szabó	Krisztina	1992-03-28
EEP-681	Kovács	Georgina	1939-08-24
NNC-477	Bács	Júlia	1935-02-05

< [] >

Advanced... Cancel < Back Next > Finish

4. Defining primary key

- ☒ Let Access add primary key.
- ☐ Choose my own primary key.
- ☐ No primary key.

OPTIONS TO IMPORT

A: Excel source, cleaning and transforming data in Excel, import into a new Access table

B: Excel source, cleaning and transforming data in Excel, import into existing Access table

C: CSV source and import into Access

D: Excel source, import, transforming and filtering data in Access

- access table: practice_data

TRANSFORM DATA – ACCESS

Transform data by query

practice_data	
*	
ID	
plate1	
plate2	
fullname	
birthdate	

Data conversion can be done in the same way as in Excel by using built-in Access functions

Functions for splitting data:

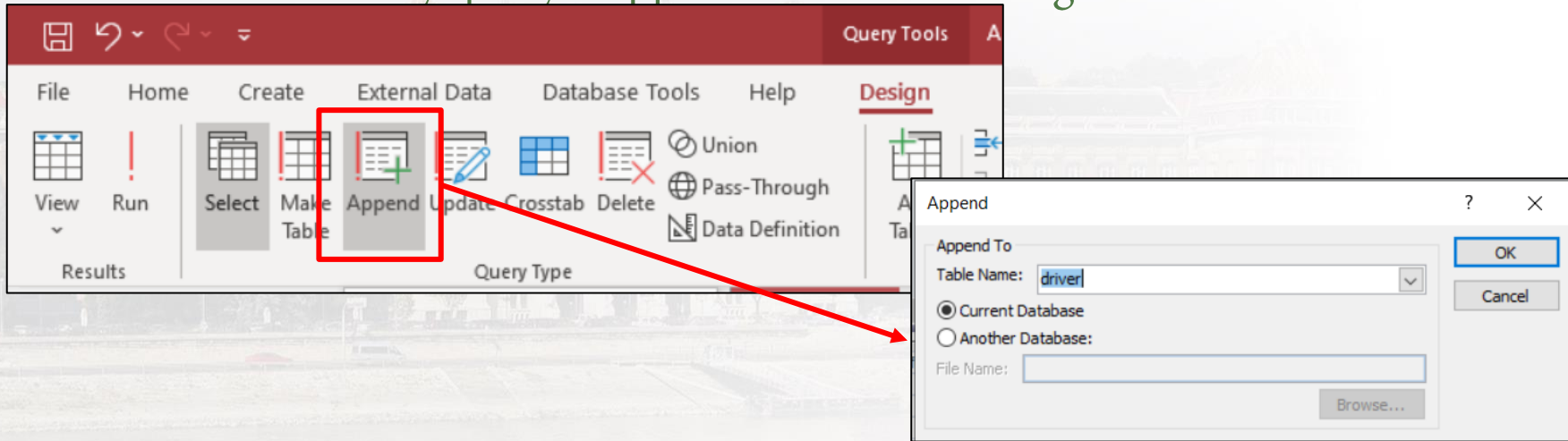
Left, Right, InStr, Len

Concatenate platenumber field1 &"-"& field2

Field:	lastname: Left([fullname];InStr([fullname];" ")-1)	firstname: Right([fullname];Len([fullname])-InStr([fullname];" "))	platenumber: [plate1] & "-" & [plate2]	birthdate
Table:				practice_data
Sort:				
Show:	☒	☒	☒	☒
Criteria:				
or:				

TRANSFORM DATA - ACCESS

Transform data by query – append result to existing table

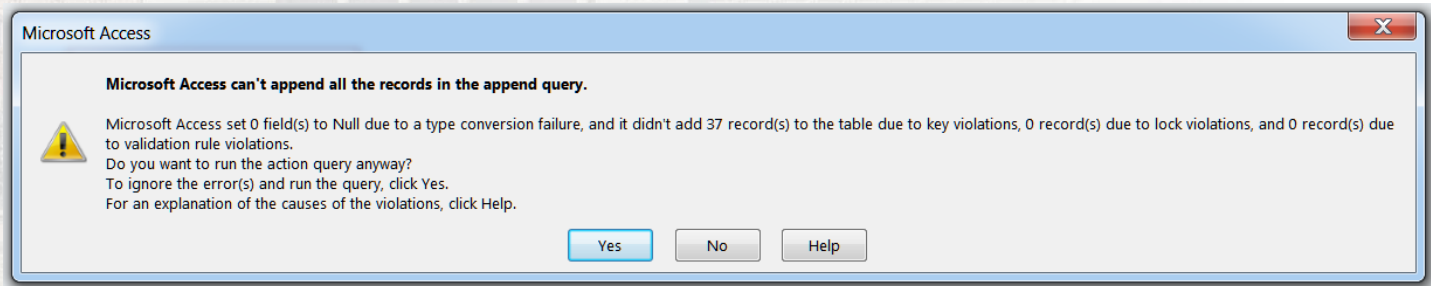


Field:	lastname: Left([fullname];InStr([fullname];" ")-1)	firstname: Right([fullname];Len([fullname])-InStr([fullname];" "))	platenumber: [plate1] & "-" & [plate2]	birthdate
Table:				practice_data
Sort:				
Append To:	lastname	firstname	platenumber	birthdate
Criteria:				
or:				

TRANSFORM DATA - ACCESS

Transform data by query – append result to existing table

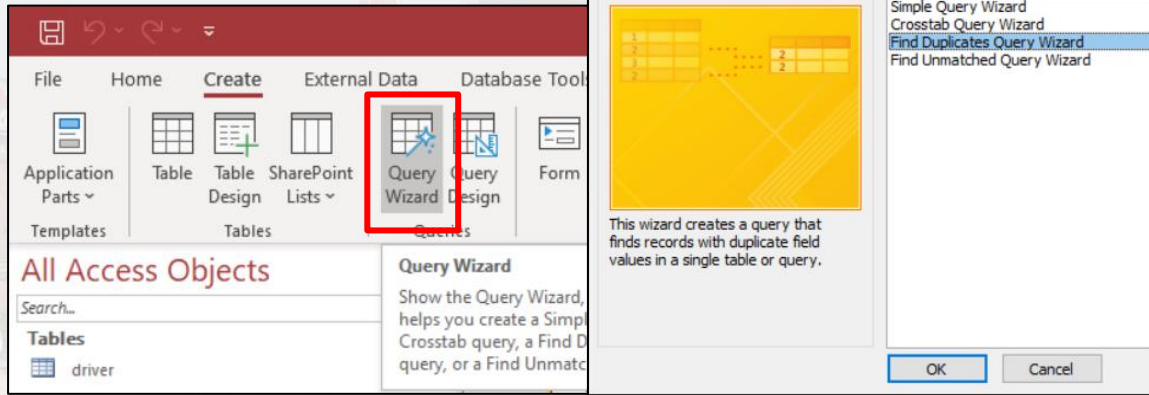
If inserted data includes primary key and it's conflicts with an existing record, warning is raised.



Field:	lastname: Left([fullname];InStr([fullname];" ")-1)	▼	firstname: Right([fullname];Len([fullname])-InStr([fullname];" "))	platenumber: [plate1] & "-" & [plate2]	birthdate
Table:					practice_data
Sort:					
Append To:	lastname		firstname	platenumber	birthdate
Criteria:					
or:					

TRANSFORM DATA - ACCESS

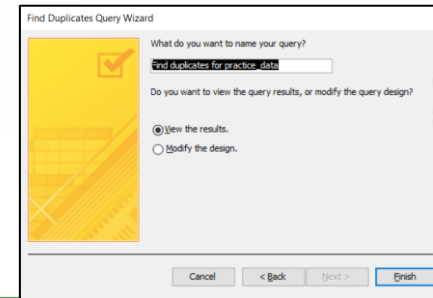
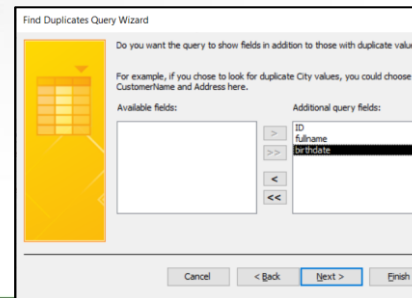
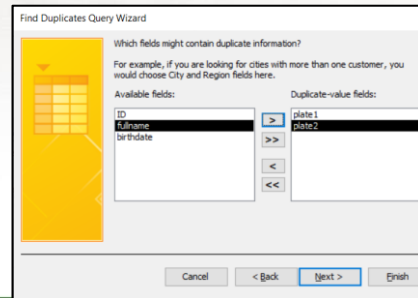
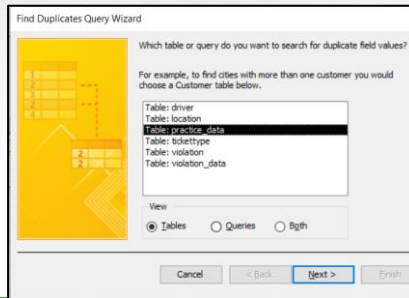
Find duplicates



Search options

- In a table
- In a query result
- Both

Potential error: if grouped by primary key all records are returned; thus every primary key differs



UPLOAD AND TRANSFORM DATA - ACCESS

Common errors

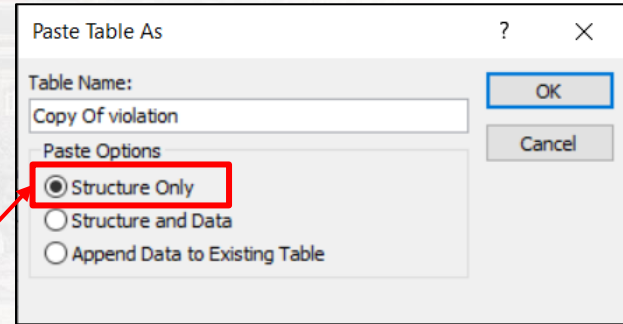
- importing data into existing table fails if field names do not agree in Excel and Access
- there is any text character (e.g. space) in a numerical column
- mixing dot, coma, semicolon
- duplicated decimal symbol
- invalid data format e.g. 12/32/2014 – there is no 32th day in December.
- multi-line header
- non-matching structural values, e.g. N/A vs NULL or parentheses
- cursor location when saving Excel file

UPLOAD AND TRANSFORM DATA - ACCESS

Revert AutoNumber

Each value is used only once, there is no possibility to restore it if deleted. There is a problem referring to shifted unique IDs.

Solution: Create a blank copy-paste table while retaining the structure, and then populate it with the correct data. The next step is to delete the old table and rename the new table and then define the connections.



UPLOAD AND TRANSFORM DATA- ACCESS

Data format imported as text (Short text) :

Access function:

FormatDateTime (*date*; [*NamedFormat*])

0-general date

1-long date

2-short date

3-long time

4-short time

Returns a formatted date or time string.

Access function:

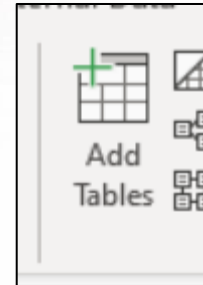
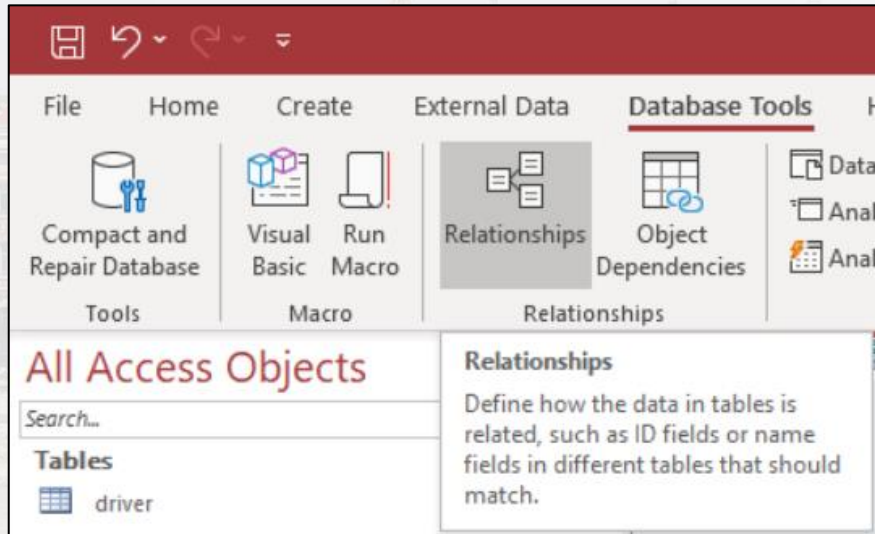
Format([table.fieldname];" ")

" " options: General Date, Long Date, Short Date, Long Time, Short Time

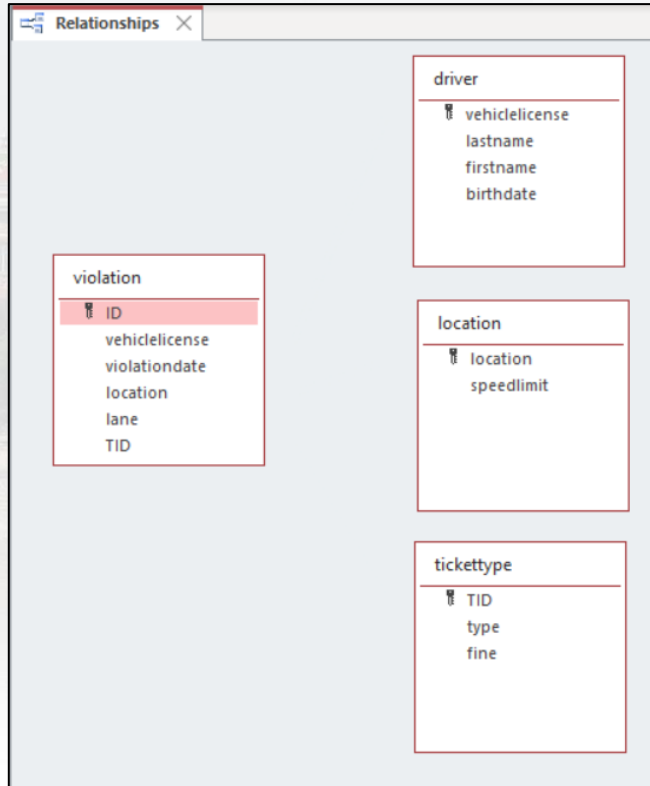
DEFINING RELATIONSHIPS OF TABLES

- Database Tools – Relationships

You need to select the tables you want to make the connections, you can add new tables with the Show Table icon

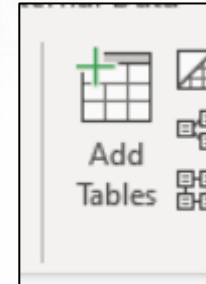


DEFINING RELATIONSHIPS OF TABLES



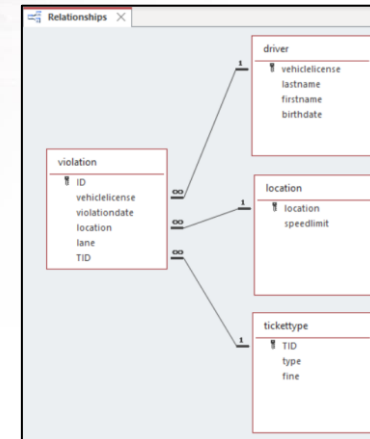
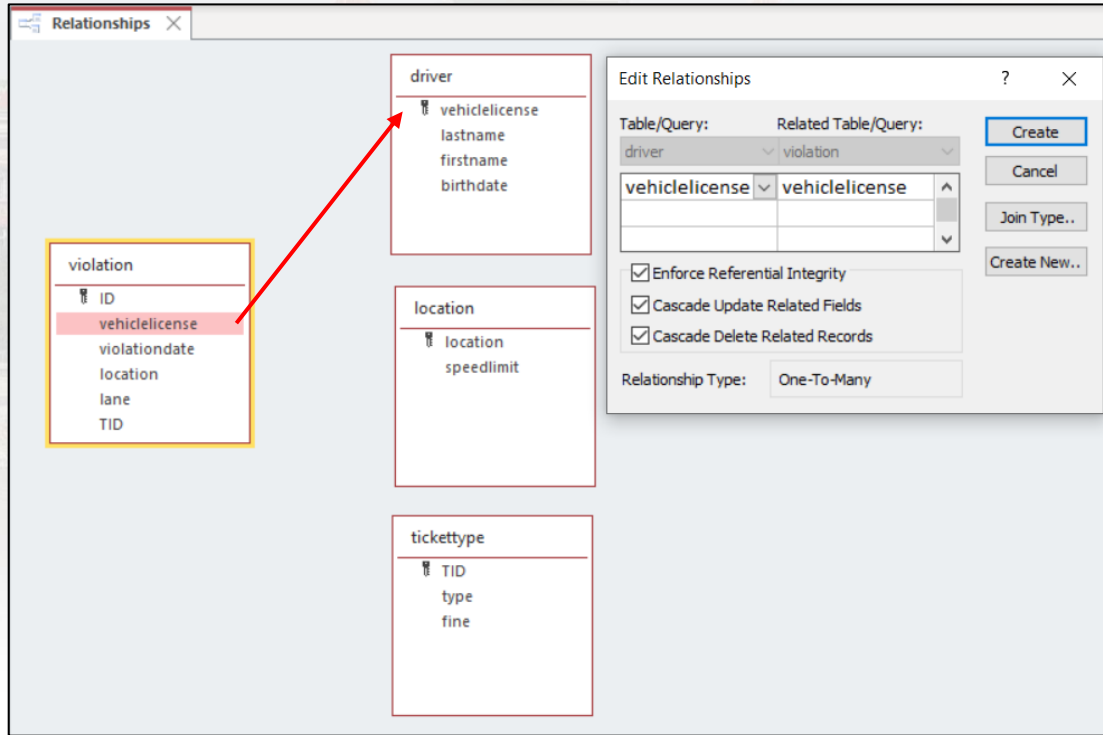
- Database Tools – Relationships

You need to select the tables you want to make the connections, you can add new tables with the Show Table icon



DEFINING RELATIONSHIPS – DRAG & DROP

- To create a Drag & Drop connection, grab one of the related fields and drag it to the other related field.



CONCLUSION

Example: Catalogue of traffic rule violations

- Defining table
- Defining relationships of tables
- Cleaning and transforming data
- Upload data and remove duplicates



Thank you for your attention!

Questions?



**BUDAPESTI MŰSZAKI
ÉS GAZDASÁGTUDOMÁNYI EGYETEM**

Építőmérnöki Kar - építőmérnöki képzés 1782 óta

Fotogrammetria és Térinformatika Tanszék