

Dealing with Extraordinary Electronic Files



“Putting a puzzle together”

Presented by:

New York State Department of Taxation and Finance

Dave Terpening, CISA, EDP Audit Specialist

Yvonne Beyer-Teal, Program Manager



True or False

- All electronic records come in Row and Column Format
- **FALSE!**
- **Beware!!! Not all electronic data files are created equal!**



Why?

- Blame the Customer?
 - Lack of experience extracting data
 - Comfort level
 - Reluctance to spend money for DP consulting
 - Legacy data
 - Application capabilities (or lack thereof)
 - Scope of and/or data requirements of audit
 - More “back-office” applications available



Tips

- Ask for a standard file format
 - direct extract from data sources
- Ask for file layouts
 - record and field definitions
- Develop a relationship with IT personnel
- Always have something to reconcile back to
- Look for patterns in data
- What's the turnaround time?



Working with these file types

“Upside ” 

- May be sources of “locked-in” unchangeable data
- May be used to compare with “normal” data received
- May be the only source of data
- Saving the customer time/money
- Gives you options



Working with these file types

“Downside ”



- May take lots of time
- You may miss something
- Steep learning curve for software
- Inexperienced staff
- Reluctance from customer to offer this data
- May need to work on site
- Compatibility issues / Familiarity with...
 - software, hardware, operating system



Some Types of Extraordinary Electronic files

- Reports stored in electronic format
 - complex arrays of data
 - multiple record types
- Adobe PDF
- “Broken-apart” transactions
 - caused by appending multiple files
- Conversion from one data format to another
- Scanned paper documents
 - NEVER A GOOD THING!!!



You ask How?

- By using “off-the-shelf”, “out-of-the-box” software!
- Understanding/Visualizing the data
 - There’s always a pattern
- Patience/Confidence
- Having something to reconcile your work to
- Second pair of “eyes”
- Ask a lot of questions
- Experiment



Examples today

- De-constructing a report file (text file) created by Vertex (using ACL)
- “Scraping” text from a PDF file (using Monarch)
- “Splicing” a delimited file containing “broken apart” transactions (using ACL)
- Maintaining data integrity converting an Excel to a dBase file (using Excel and dBase)
- Eliminating accounting washing entries (Using ACL)



Warning !!

This is not a perfect science !!

The "Array" (Vertex)

```

ugly_file.txt - Notepad
File Edit Format View Help

RELEASE: 1.3 00/00/0000
RUN DATE 01/01/2001

                                XXXXX XXX REGISTER - DETAIL XXX
                                TRANS PERIOD:      START - END
                                INVOICE PERIOD: MM/DD/YYYY - MM/DD/YYYY
COMPANY-CD: ALL  DIVISION-CD: ALL  TYPE: ALL          SUBTYPE: ALL          TAX-TYPE: ALL  STATE: XXXXXXXX
                                PAGE 1

-----
STATE              COUNTY              CITY              DISTRICT              TOTAL TAX
-----
.....XXXXXXXXX
-----
CITY: 0001          COUNTY: 0002      GEOCODE: 33-001-0000 STR:          T 33-001-0000 F 33-001-0000 A 33-001-0000
COMPANY-CD: 0400  DIVISION-CD:      CUST-CD: 0000011112          TRANS-CD: X  TYPE: XXXX  SUBTYPE: 1
EXEMPT CERTIFICATE #:
INV: 0090038654  DATE: 11/02/1997  LINE: 00001  PROD-CD:          USER:
TAX RATE/TYPE----- .040000 S          .040000 S          0.000000 X          0.000000 X
TAXABLE AMT----- -2,023.75          -2,023.75          -2,023.75          -2,023.75
TAX AMT----- -80.95          0.00          0.00          0.00
CITY: 0001          COUNTY: 0002      GEOCODE: 33-001-0000 STR:          T 33-001-0000 F 33-001-0000 A 33-001-0000
COMPANY-CD: 0400  DIVISION-CD:      CUST-CD: 0000011112          TRANS-CD: X  TYPE: XXXX  SUBTYPE: 1
EXEMPT CERTIFICATE #:
INV: 0090038655  DATE: 11/02/1997  LINE: 00001  PROD-CD:          USER:
TAX RATE/TYPE----- .040000 S          .040000 S          0.000000 X          0.000000 X
TAXABLE AMT----- -1,953.75          -1,953.75          -1,953.75          -1,953.75
TAX AMT----- -78.15          0.00          0.00          0.00
CITY: 0001          COUNTY: 0002      GEOCODE: 33-001-0000 STR:          T 33-001-0000 F 33-001-0000 A 33-001-0000
COMPANY-CD: 0400  DIVISION-CD:      CUST-CD: 0000011112          TRANS-CD: X  TYPE: XXXX  SUBTYPE: 1
EXEMPT CERTIFICATE #:
INV: 0090038656  DATE: 11/02/1997  LINE: 00001  PROD-CD:          USER:
TAX RATE/TYPE----- .040000 S          .040000 S          0.000000 X          0.000000 X
TAXABLE AMT----- -1,056.25          -1,056.25          -1,056.25          -1,056.25
TAX AMT----- -42.25          0.00          0.00          0.00
CITY: 0001          COUNTY: 0002      GEOCODE: 33-001-0000 STR:          T 33-001-0000 F 33-001-0000 A 33-001-0000
COMPANY-CD: 0400  DIVISION-CD:      CUST-CD: 0000011112          TRANS-CD: X  TYPE: XXXX  SUBTYPE: 1
EXEMPT CERTIFICATE #:
INV: 0090038657  DATE: 11/02/1997  LINE: 00001  PROD-CD:          USER:

```

The Adobe PDF file



Trust US Rapid Oil Change Company

"We use nothing but the best used oil!"



Store: AQ5001

Invoice Date: 06/02/2004

September 6, 2007

5:31:10 PM

Page 1 of 84

Invoice	Type	ID	Customer	Time	Gross Rev	Coupon	Discnt	Allow	Amount	Tax Amt	Total	Payment
116165	R	CRA45N39Y	RFIFENRIFENBURNBU	08:51:23	41.78	R5			5.00	2.80	36.78	MasterCard
116166	R	BNM52N37Y	RAHANTSHANTLERTLE	08:56:24	33.01	NEW			0.00	2.52	33.01	Visa
116167	T	CBX93N46Y	RURNEBURNESES	09:49:43	0.00				0.00	0.00	0.00	
116168	T	ACL37N20Y	TCUCHEBUCHHOLTHOL	09:55:11	0.00				0.00	0.00	0.00	
116169	R	APF77N91Y	ZAZZAHAZZAMAN	10:08:47	30.89	FFBA			6.00	1.90	24.89	Visa
116170	R	APW87N57Y	NANGEGANGEMIEMI	10:32:25	30.89	FFBA	RBT		6.00	1.90	24.89	MasterCard
116171	R	ASA69N62Y	ESASBIVASBINDEIND	10:34:29	30.89	FFBA	RBT		6.00	1.90	24.89	Cash
116172	R	CHT77N63Y	EHEEMTHEEMANMAN	10:57:46	42.62	CCLD	RBT		8.00	2.64	34.62	MasterCard
116173	R	A877KNAY	ARANKFRANKEKE	11:16:02	30.72	CCLD	MISC		8.00	1.73	22.72	Visa
116174	R	BLN94N11Y	AWALDAWALDD	11:24:39	30.72	CCLD	RBT		8.00	1.73	22.72	Cash
116175	R	EGB42N9Y	SASSYTASSY	11:37:57	58.43	RET			0.00	4.45	58.43	Cash
116176	R	AW337N56Y	ILINKKLINKKOKO	11:44:14	30.89	FFDD	RBT		6.00	1.90	24.89	Visa
116177	R	AYJ94N47Y	TSASHIWASHINGTING	12:12:22	30.89	FFBA	NEW		6.00	1.90	24.89	Cash
116178	R	BLM34N52Y	IHISSWHISSELSSEL	12:21:28	31.38	RET			0.00	2.39	31.38	Visa
116179	R	APC35N32Y	IEISWEIS	12:33:29	39.99	GS90			39.99	0.00	0.00	
116180	R	CTJ74N40Y	TATYBATY	12:45:03	31.38	NEW			0.00	2.39	31.38	Cash
116181	R	BA632N7Y	SUSECBUSECKCK	12:53:37	34.22	MISC	CCL		5.00	2.23	29.22	Visa
116182	R	BWV60N39Y	RIRONTIRONENE	13:40:31	31.22	AAA2	RBT		2.00	2.23	29.22	MasterCard
116183	R	CTJ10N35Y	N'NEIO'NBILIL	13:56:04	36.13	CCLD	RBT		8.00	2.14	28.13	MasterCard
116184	R	AVG54N69Y	CJOJTOJOWJTOJOWCOWI	14:20:22	51.94	RET			0.00	3.96	51.94	Cash
116185	R	BVN75N08Y	RURKBBURKEE	14:42:29	77.91	RET			0.00	5.94	77.91	Cash
116186	R	BXG33N08Y	EIERIPIERII	15:05:31	81.56	CCLD			8.00	5.61	73.56	Visa
116187	R	AHB41N32Y	NINDEBINDERERT	15:07:58	58.43	RET			0.00	4.45	58.43	MasterCard
116188	R	57509NJB	NMRMSTARMSTRONTR	15:27:23	31.38	RET			0.00	2.39	31.38	MasterCard
116189	R	CSV45N50Y	NINGESINGERER	15:47:12	0.00	NEW	CCAJ		0.00	0.00	0.00	
116190	R	BEM45N02Y	LOLLAHOLLANDAND	16:14:23	83.76	CCLD			8.00	5.77	75.76	MasterCard



The "Broken-Apart" File

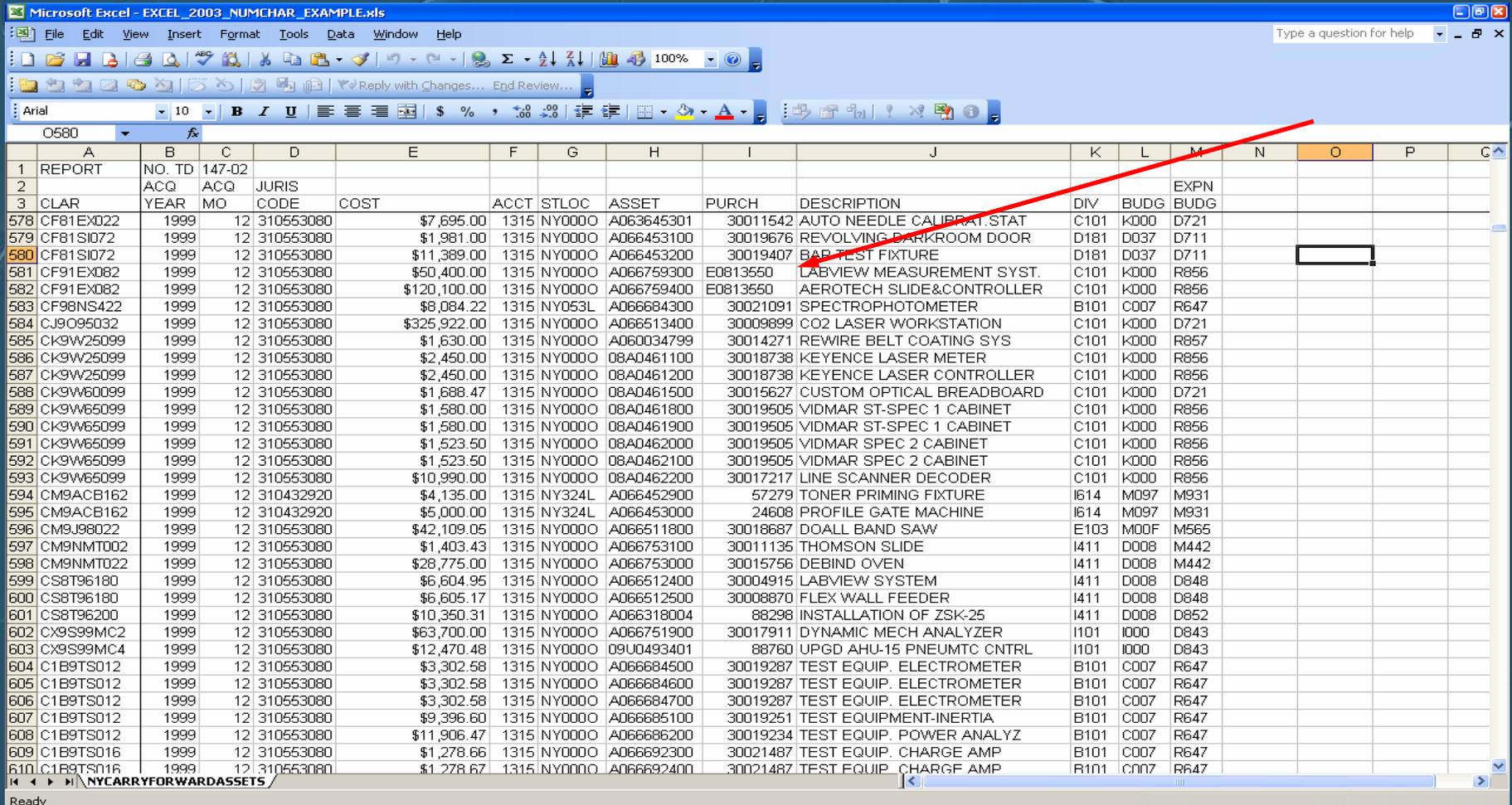
```
SPLIT_REC_EXAMP1.FIL - Notepad
File Edit Format View Help
10017F ANYTOWN CTRL SCH DIST KOA SMALLVILLE SCHOOL RTE 2 SMALLVILLE NY
10017F ANYTOWN CTRL SCH DIST KOA SMALLVILLE SCHOOL RTE 2 SMALLVILLE NY
```

```
SPLIT_REC_EXAMP1.FIL - Notepad
File Edit Format View Help
66135 2003-10-10 00:00:00.000 2 - #2 Fuel 297.89999999999998 .9160000 27.
90772 2003-11-17 00:00:00.000 2 - #2 Fuel 837.0 .9468000 792.4700 .0
```

```
SPLIT_REC_EXAMP1.FIL - Notepad
File Edit Format View Help
.0000000 .0000 Renss .0000 4000 SIGN TKBK LVE 1 COPY/DELIVER
.0000 Renss 4000 SIGN TKBK LVE 1 COPY/DELIVER BETWN 7AM & 6PM
```

Excel to dBase

"Original Excel file"



Microsoft Excel - EXCEL_2003_NUMCHAR_EXAMPLE.xls

Type a question for help

0580

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	C
1	REPORT	NO. TD	147-02														
2		ACQ	ACQ	JURIS													
3	CLAR	YEAR	MO	CODE	COST	ACCT	STLOC	ASSET	PURCH	DESCRIPTION	DIV	BUDG	BUDG				
578	CF81EX022	1999	12	310553080	\$7,695.00	1315	NY0000	A063645301	30011542	AUTO NEEDLE CALIBRAT.STAT	C101	K000	D721				
579	CF81S1072	1999	12	310553080	\$1,981.00	1315	NY0000	A066453100	30019676	REVOLVING DARKROOM DOOR	D181	D037	D711				
580	CF81S1072	1999	12	310553080	\$11,389.00	1315	NY0000	A066453200	30019407	BAR TEST FIXTURE	D181	D037	D711				
581	CF91EX082	1999	12	310553080	\$50,400.00	1315	NY0000	A066759300	E0813550	LABVIEW MEASUREMENT SYST.	C101	K000	R856				
582	CF91EX082	1999	12	310553080	\$120,100.00	1315	NY0000	A066759400	E0813550	AEROTECH SLIDE&CONTROLLER	C101	K000	R856				
583	CF98NS422	1999	12	310553080	\$8,084.22	1315	NY053L	A066684300	30021091	SPECTROPHOTOMETER	B101	C007	R647				
584	CJ9095032	1999	12	310553080	\$325,922.00	1315	NY0000	A066513400	30009899	CO2 LASER WORKSTATION	C101	K000	D721				
585	CK9W25099	1999	12	310553080	\$1,630.00	1315	NY0000	A060034799	30014271	REWIRE BELT COATING SYS	C101	K000	R857				
586	CK9W25099	1999	12	310553080	\$2,450.00	1315	NY0000	08A0461100	30018738	KEYENCE LASER METER	C101	K000	R856				
587	CK9W25099	1999	12	310553080	\$2,450.00	1315	NY0000	08A0461200	30018738	KEYENCE LASER CONTROLLER	C101	K000	R856				
588	CK9W60099	1999	12	310553080	\$1,688.47	1315	NY0000	08A0461500	30015627	CUSTOM OPTICAL BREADBOARD	C101	K000	D721				
589	CK9W65099	1999	12	310553080	\$1,580.00	1315	NY0000	08A0461800	30019505	VIDMAR ST-SPEC 1 CABINET	C101	K000	R856				
590	CK9W65099	1999	12	310553080	\$1,580.00	1315	NY0000	08A0461900	30019505	VIDMAR ST-SPEC 1 CABINET	C101	K000	R856				
591	CK9W65099	1999	12	310553080	\$1,523.50	1315	NY0000	08A0462000	30019505	VIDMAR SPEC 2 CABINET	C101	K000	R856				
592	CK9W65099	1999	12	310553080	\$1,523.50	1315	NY0000	08A0462100	30019505	VIDMAR SPEC 2 CABINET	C101	K000	R856				
593	CK9W65099	1999	12	310553080	\$10,990.00	1315	NY0000	08A0462200	30017217	LINE SCANNER DECODER	C101	K000	R856				
594	CM9ACB162	1999	12	310432920	\$4,135.00	1315	NY324L	A066452900	57279	TONER PRIMING FIXTURE	I614	M097	M931				
595	CM9ACB162	1999	12	310432920	\$5,000.00	1315	NY324L	A066453000	24608	PROFILE GATE MACHINE	I614	M097	M931				
596	CM9J98022	1999	12	310553080	\$42,109.05	1315	NY0000	A066511800	30018687	DOALL BAND SAW	E103	M00F	M565				
597	CM9NMT002	1999	12	310553080	\$1,403.43	1315	NY0000	A066753100	30011135	THOMSON SLIDE	I411	D008	M442				
598	CM9NMT022	1999	12	310553080	\$28,775.00	1315	NY0000	A066753000	30015756	DEBIND OVEN	I411	D008	M442				
599	CS8T96180	1999	12	310553080	\$6,604.95	1315	NY0000	A066512400	30004915	LABVIEW SYSTEM	I411	D008	D848				
600	CS8T96180	1999	12	310553080	\$6,605.17	1315	NY0000	A066512500	30008870	FLEX WALL FEEDER	I411	D008	D848				
601	CS8T96200	1999	12	310553080	\$10,350.31	1315	NY0000	A066318004	88298	INSTALLATION OF ZSIK-25	I411	D008	D852				
602	CX9S99MC2	1999	12	310553080	\$63,700.00	1315	NY0000	A066751900	30017911	DYNAMIC MECH ANALYZER	I101	I000	D843				
603	CX9S99MC4	1999	12	310553080	\$12,470.48	1315	NY0000	09U0493401	88760	UPGD AHU-15 PNEUMTC CNTRL	I101	I000	D843				
604	C1B9TS012	1999	12	310553080	\$3,302.58	1315	NY0000	A066684500	30019287	TEST EQUIP. ELECTROMETER	B101	C007	R647				
605	C1B9TS012	1999	12	310553080	\$3,302.58	1315	NY0000	A066684600	30019287	TEST EQUIP. ELECTROMETER	B101	C007	R647				
606	C1B9TS012	1999	12	310553080	\$3,302.58	1315	NY0000	A066684700	30019287	TEST EQUIP. ELECTROMETER	B101	C007	R647				
607	C1B9TS012	1999	12	310553080	\$9,396.60	1315	NY0000	A066685100	30019251	TEST EQUIPMENT-INERTIA	B101	C007	R647				
608	C1B9TS012	1999	12	310553080	\$11,906.47	1315	NY0000	A066686200	30019234	TEST EQUIP. POWER ANALYZ	B101	C007	R647				
609	C1B9TS016	1999	12	310553080	\$1,278.66	1315	NY0000	A066692300	30021487	TEST EQUIP. CHARGE AMP	B101	C007	R647				
610	C1B9TS016	1999	12	310553080	\$1,278.67	1315	NY0000	A066692400	30021487	TEST EQUIP. CHARGE AMP	B101	C007	R647				

NYCARRYFORWARDASSETS

Ready

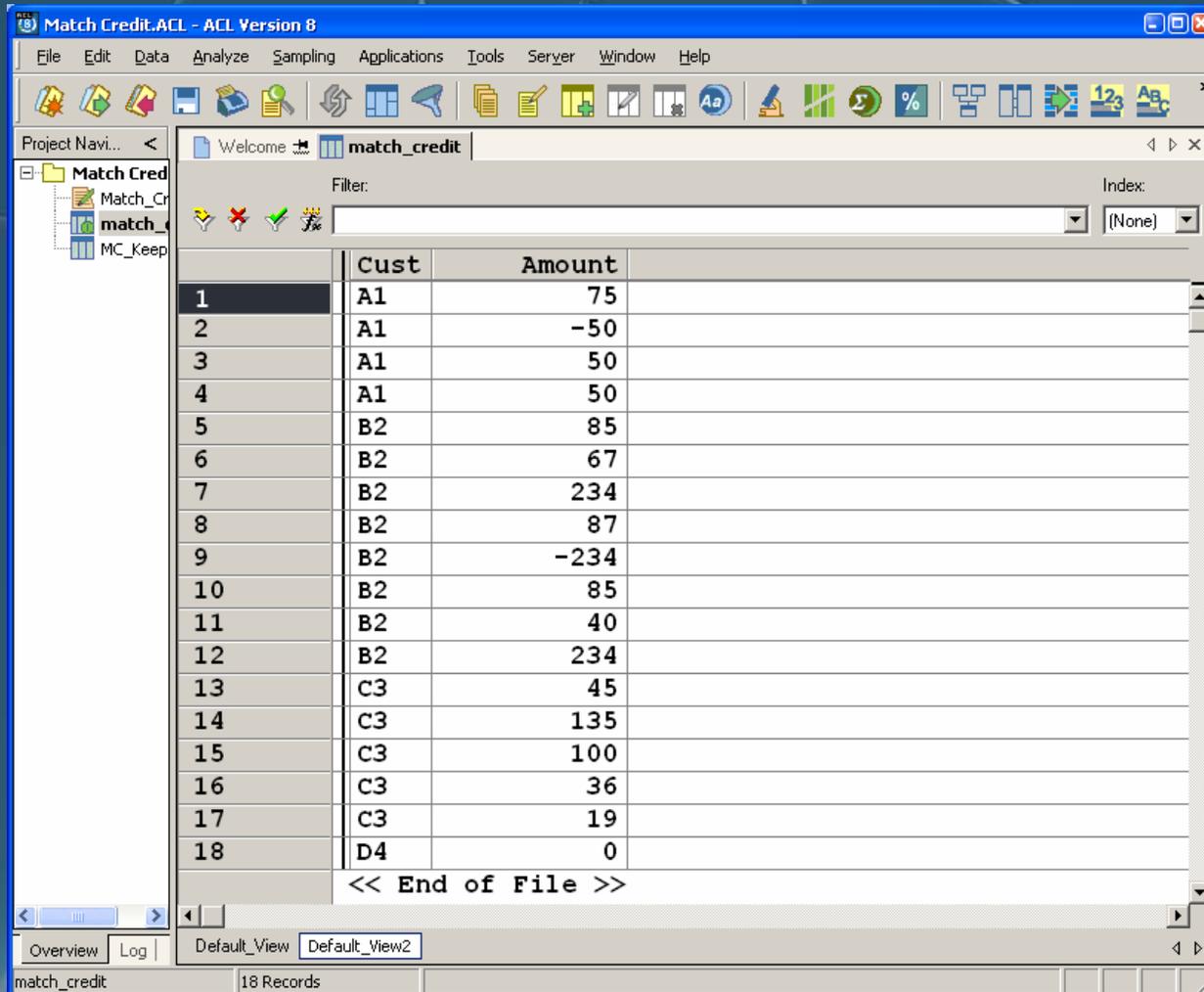
Excel to dBase

"Incorrectly converted from Excel to dBase"

Excel_2003_numchar_example

Clar	Year	Mo	Code	Cost	Acct	Stloc	Asset	Purch	Descriptio	Div	Budg	Budg
CF81EX022	1999	12	310553080	7695.00	1315	NY0000	A063645301	30011542	AUTO NEEDLE CALIBRAT STAT	C101	K000	D721
CF81SI072	1999	12	310553080	1981.00	1315	NY0000	A066453100	30019676	REVOLVING DARKROOM DOOR	D181	D037	D711
CF81SI072	1999	12	310553080	11389.00	1315	NY0000	A066453200	30019407	BAR TEST FIXTURE	D181	D037	D711
CF91EX082	1999	12	310553080	50400.00	1315	NY0000	A066759300		LABVIEW MEASUREMENT SYST.	C101	K000	R856
CF91EX082	1999	12	310553080	120100.00	1315	NY0000	A066759400		AEROTECH SLIDE&CONTROLLER	C101	K000	R856
CF98NS422	1999	12	310553080	8084.22	1315	NY053L	A066684300	30021091	SPECTROPHOTOMETER	B101	C007	R647
CJ9095032	1999	12	310553080	325922.00	1315	NY0000	A066513400	30009899	CO2 LASER WORKSTATION	C101	K000	D721
CK9W25099	1999	12	310553080	1630.00	1315	NY0000	A060034799	30014271	REWIRE BELT COATING SYS	C101	K000	R857
CK9W25099	1999	12	310553080	2450.00	1315	NY0000	08A0461100	30018738	KEYENCE LASER METER	C101	K000	R856
CK9W25099	1999	12	310553080	2450.00	1315	NY0000	08A0461200	30018738	KEYENCE LASER CONTROLLER	C101	K000	R856
CK9W60099	1999	12	310553080	1688.47	1315	NY0000	08A0461500	30015627	CUSTOM OPTICAL BREADBOARD	C101	K000	D721
CK9W65099	1999	12	310553080	1580.00	1315	NY0000	08A0461800	30019505	VIDMAR ST-SPEC 1 CABINET	C101	K000	R856
CK9W65099	1999	12	310553080	1580.00	1315	NY0000	08A0461900	30019505	VIDMAR ST-SPEC 1 CABINET	C101	K000	R856
CK9W65099	1999	12	310553080	1523.50	1315	NY0000	08A0462000	30019505	VIDMAR SPEC 2 CABINET	C101	K000	R856
CK9W65099	1999	12	310553080	1523.50	1315	NY0000	08A0462100	30019505	VIDMAR SPEC 2 CABINET	C101	K000	R856
CK9W65099	1999	12	310553080	10990.00	1315	NY0000	08A0462200	30017217	LINE SCANNER DECODER	C101	K000	R856
CM9ACB162	1999	12	310432920	4135.00	1315	NY324L	A066452900	57279	TONER PRIMING FIXTURE	I614	M097	M931
CM9ACB162	1999	12	310432920	5000.00	1315	NY324L	A066453000	24608	PROFILE GATE MACHINE	I614	M097	M931
CM9J98022	1999	12	310553080	42109.05	1315	NY0000	A066511800	30018687	DOALL BAND SAW	E103	M00F	M565
CM9NMT002	1999	12	310553080	1403.43	1315	NY0000	A066753100	30011135	THOMSON SLIDE	I411	D008	M442
CM9NMT022	1999	12	310553080	28775.00	1315	NY0000	A066753000	30015756	DEBIND OVEN	I411	D008	M442
CS8T96180	1999	12	310553080	6604.95	1315	NY0000	A066512400	30004915	LABVIEW SYSTEM	I411	D008	D848
CS8T96180	1999	12	310553080	6605.17	1315	NY0000	A066512500	30008870	FLEX WALL FEEDER	I411	D008	D848
CS8T96200	1999	12	310553080	10350.31	1315	NY0000	A066318004	88298	INSTALLATION OF ZSK-25	I411	D008	D852
CX9S99MC2	1999	12	310553080	63700.00	1315	NY0000	A066751900	30017911	DYNAMIC MECH ANALYZER	I101	I000	D843
CX9S99MC4	1999	12	310553080	12470.48	1315	NY0000	09U0493401	88760	UPGD AHU-15 PNEUMTC CNTRL	I101	I000	D843
C189TS012	1999	12	310553080	3302.58	1315	NY0000	A066684500	30019287	TEST EQUIP. ELECTROMETER	B101	C007	R647
C189TS012	1999	12	310553080	3302.58	1315	NY0000	A066684600	30019287	TEST EQUIP. ELECTROMETER	B101	C007	R647
C189TS012	1999	12	310553080	3302.58	1315	NY0000	A066684700	30019287	TEST EQUIP. ELECTROMETER	B101	C007	R647
C189TC012	1999	12	310553080	3302.58	1315	NY0000	A066684500	30019287	TEST EQUIPMENT IMPEDIA	B101	C007	R647

"Washing Entries"



The screenshot displays the Match Credit.ACL software interface. The main window shows a table with 18 records. The columns are labeled 'Cust' and 'Amount'. The data is as follows:

	Cust	Amount
1	A1	75
2	A1	-50
3	A1	50
4	A1	50
5	B2	85
6	B2	67
7	B2	234
8	B2	87
9	B2	-234
10	B2	85
11	B2	40
12	B2	234
13	C3	45
14	C3	135
15	C3	100
16	C3	36
17	C3	19
18	D4	0

At the bottom of the table, the text '<< End of File >>' is displayed. The software interface includes a menu bar (File, Edit, Data, Analyze, Sampling, Applications, Tools, Server, Window, Help), a toolbar with various icons, and a project navigation pane on the left showing a folder structure with 'Match Cred', 'Match_Cr', 'match_', and 'MC_Keep'.



Exceptions/Issues

- Proprietary file formats
 - specific to software application package
- Encrypted data
- “Imaged” data
- Corrupted data
- Can't afford to buy the software
- No support from
 - audit customer
 - audit staff
 - management



Contacts

Yvonne Beyer-Teal, Program Manager, Technology Assist Audits
Section

New York State Department of Taxation and Finance

Email: Yvonne_Beyer-Teal@tax.state.ny.us

Dave Terpening, CISA

Technology Assist Audits Section

Technology Integration Team

New York State Department of Taxation and Finance

Email: David_Terpening@tax.state.ny.us



Questions ?

Thank You !