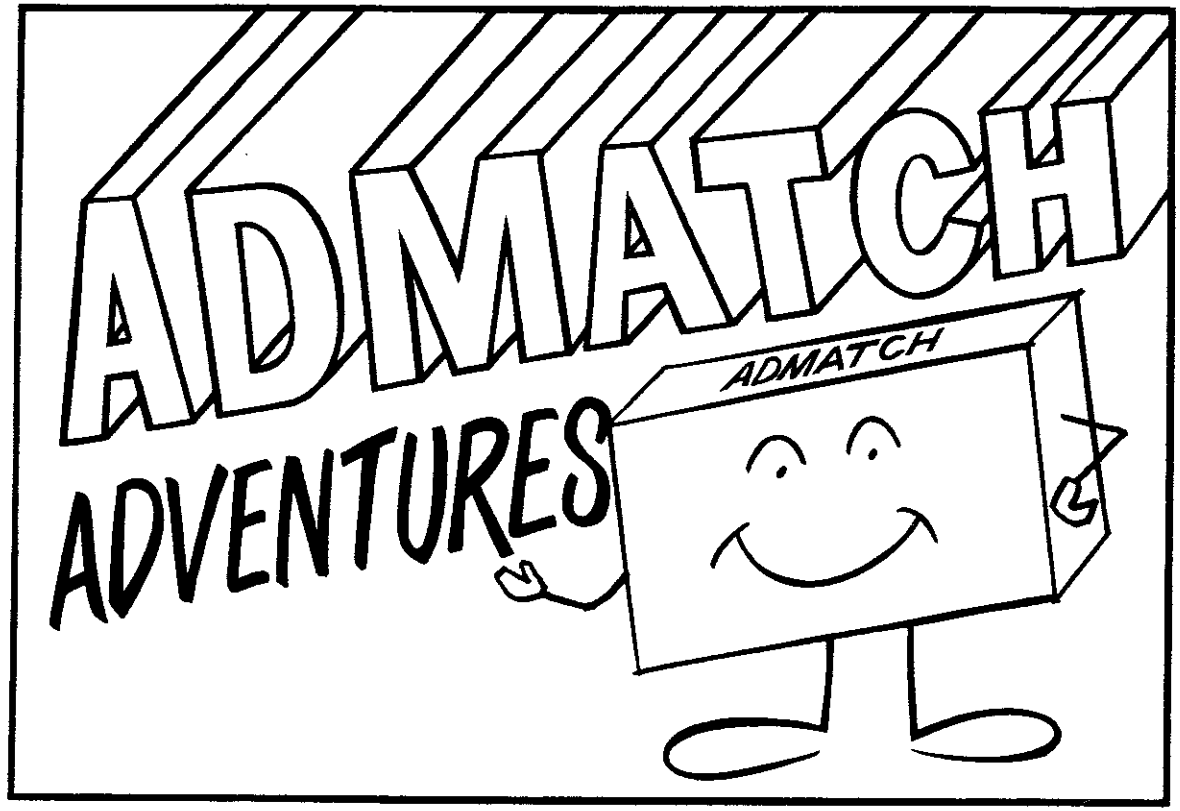
census use study

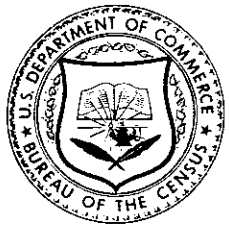


by Donald F. Cooke
and CUS staff

U. S. DEPARTMENT OF COMMERCE
Frederick B. Dent, Secretary
Sidney L. Jones, Assistant Secretary
for Economic Affairs

Social and Economic Statistics Administration
Edward D. Failor, Administrator

BUREAU OF THE CENSUS
Vincent P. Barabba, Director
Robert L. Hagan, Deputy Director



BUREAU OF THE CENSUS

Vincent P. Barabba, Director
Robert L. Hagan, Deputy Director

DATA USERS SERVICES OFFICE

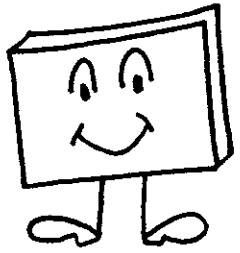
Robert B. Voight, Chief

The Census Use Study is under the direction of **Caby C. Smith,** Project Director.

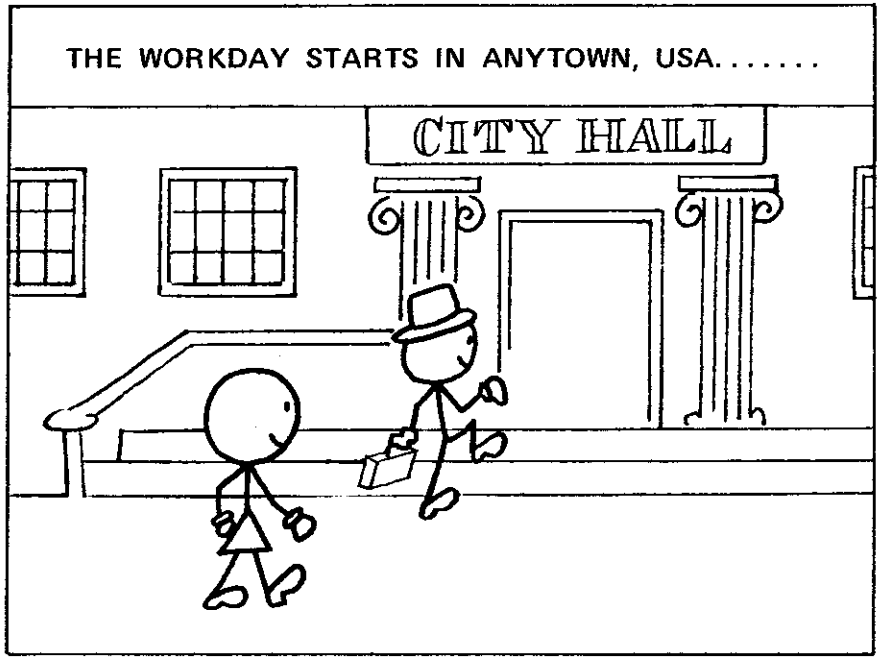
The ADMATCH Adventures comic strip was written and designed by **Donald F. Cooke** with the assistance of the entire Census Use Study staff.

It represents the beginning of a series of Census Use publications aimed toward making complex technological advancements meaningful to public administrators.

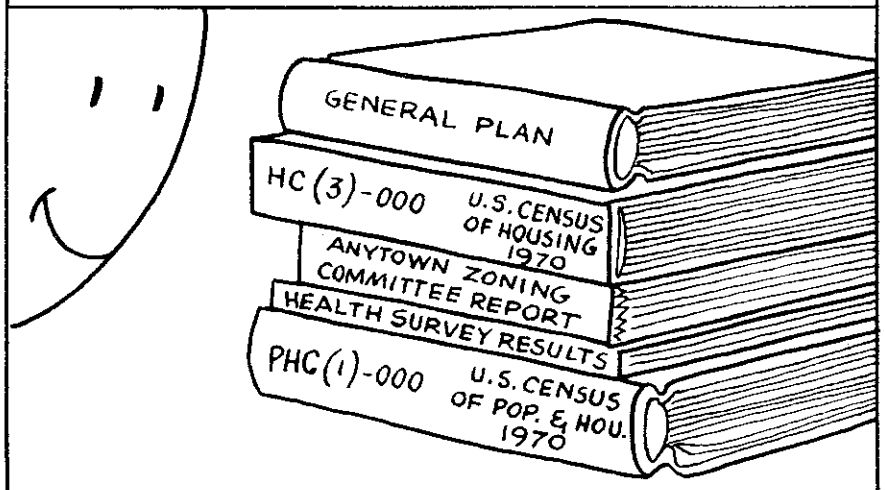
ADMATCH ADVENTURES



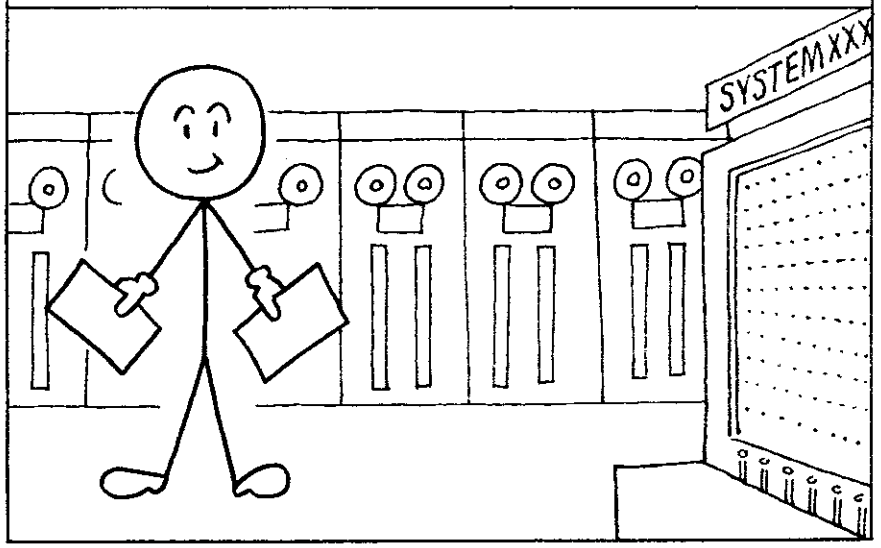
By: Donald F. Cooke



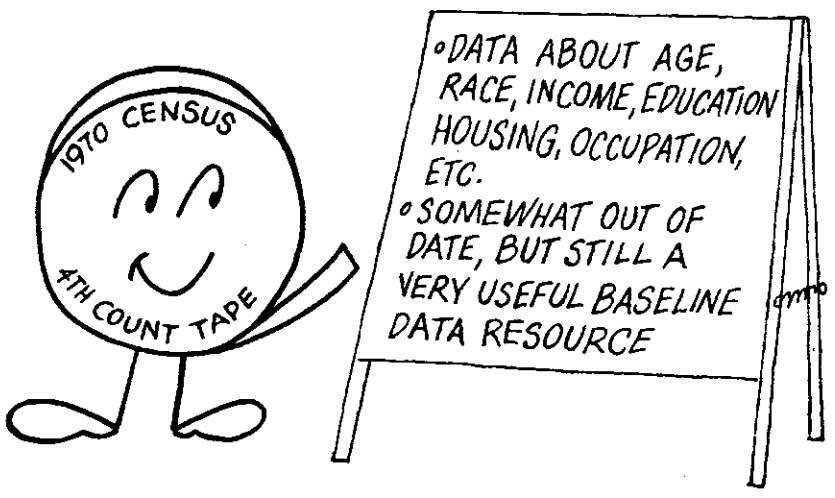
THE CITY ANALYSTS AND PLANNERS ARE FAMILIAR WITH THE INFORMATION SOURCES IN THE CITY.



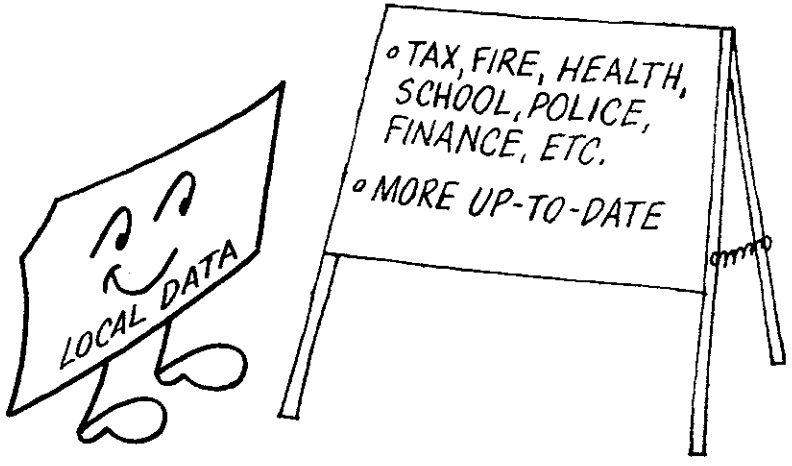
SOME HAVE EVEN USED THE DATA PROCESSING DEPARTMENT'S COMPUTER.....



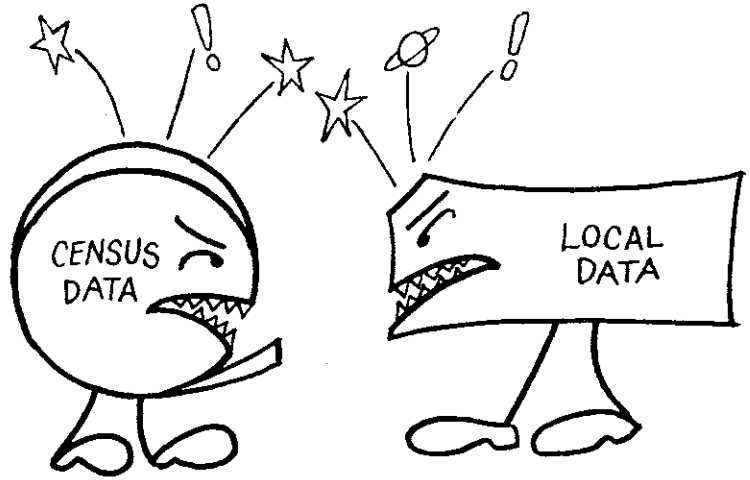
THERE IS COMPUTERIZED DATA FROM THE CENSUS BUREAU.....



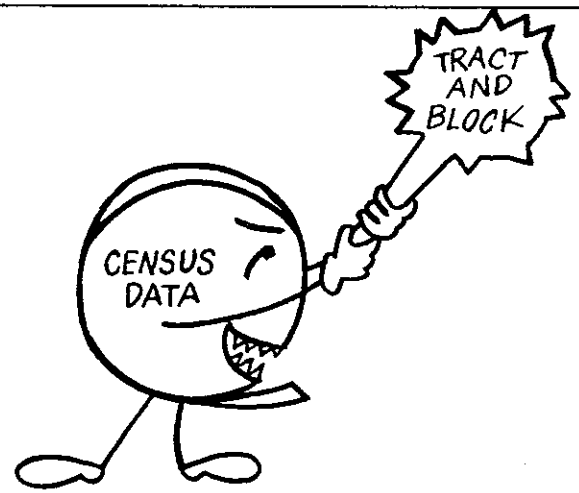
AND FROM LOCAL DATA SOURCES...



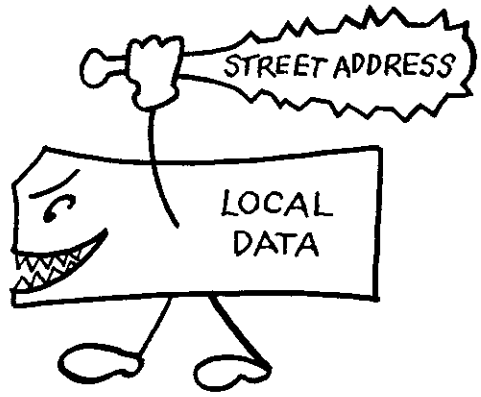
UNFORTUNATELY, THESE DATA RESOURCES DON'T AGREE ON GEOGRAPHY.



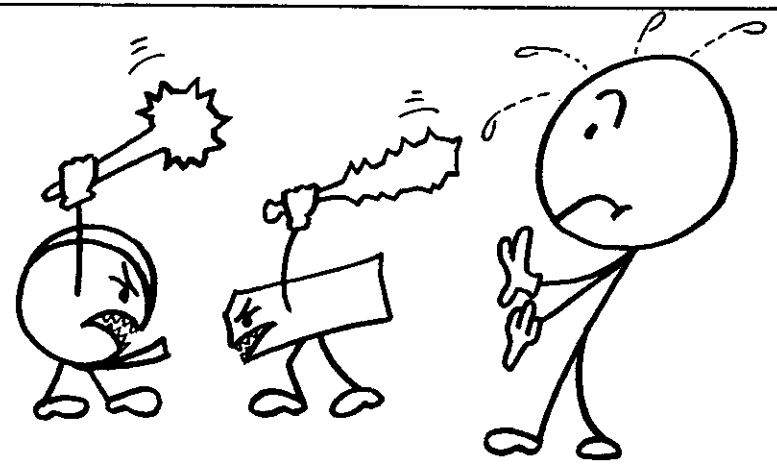
1970 CENSUS DATA TAPES COME AGGREGATED TO THE TRACT OR BLOCK LEVEL.



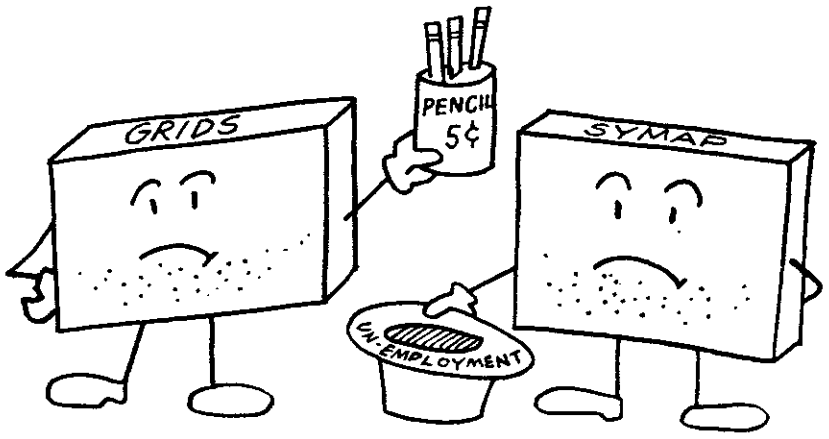
LOCAL DATA IS AGGREGATED TO INDIVIDUAL SPECIAL PURPOSE AREAS SUCH AS HEALTH DISTRICT OR NOT AGGREGATED AT ALL.



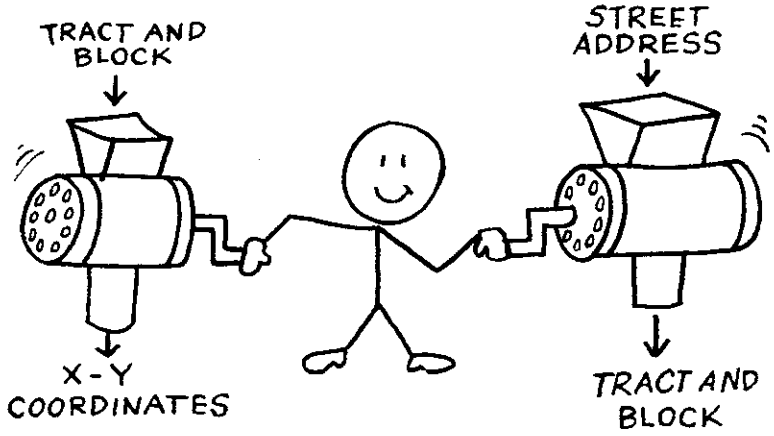
GEOGRAPHIC DISAGREEMENTS CAUSE GREAT CONSTERNATION TO POTENTIAL DATA USERS. . . .



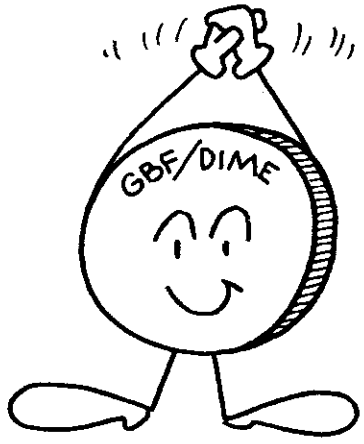
BESIDES THIS, BOTH DATA TYPES LACK GRID COORDINATES NECESSARY FOR COMPUTER MAPPING AND COORDINATE DATA RETRIEVAL.



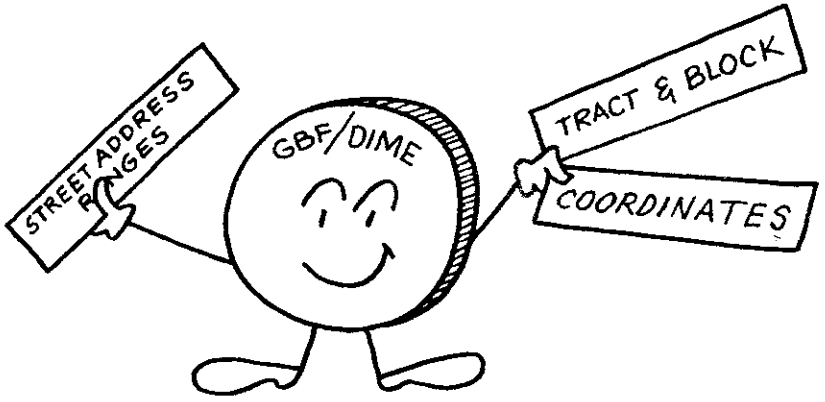
THESE PROBLEMS COULD BE SOLVED IF WE COULD TRANSLATE THE GEOCODES (GEOGRAPHIC CODES--ADDRESSES, TRACTS AND BLOCKS, COORDINATES).



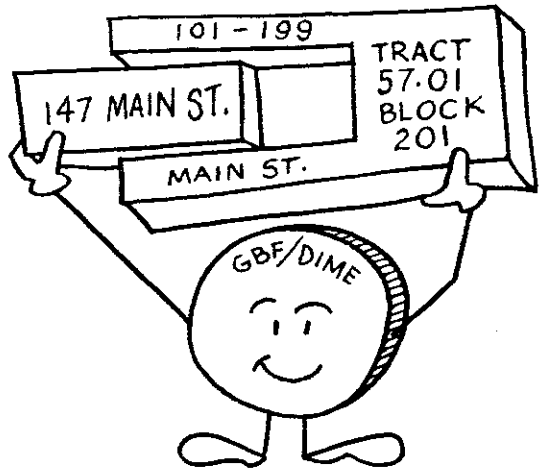
PART OF THE SOLUTION LIES IN THE GBF-DIME FILES PRODUCED BY THE CENSUS BUREAU AROUND 1970.



THEY SHOW THE RELATIONSHIP BETWEEN ADDRESS RANGES, TRACT AND BLOCK NUMBERS, AND GRID COORDINATES.



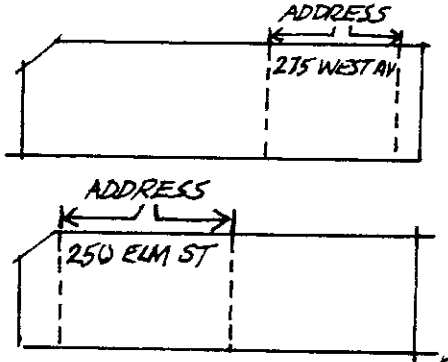
EACH ADDRESS IN THE CITY SHOULD FIT INTO A GBF-DIME FILE ADDRESS RANGE, LINKING IT TO OTHER GEOCODES.



HOWEVER (UNLIKE OUR ANIMATED DIME FILE HERE), GBF-DIME FILES ARE STATIC ENTITIES—THEY REQUIRE AN OUTSIDE PROGRAM TO LOOK UP THE BEST ADDRESS RANGE FOR A DATA RECORD.

THE ADDRESS MATCHING PROBLEM IS COMPLICATED BY:

① VARYING DATA FORMATS



② DIFFERENT ABBREVIATION CONVENTIONS:

AV.
AVE
AVENUE

③ SPELLING VARIATIONS:

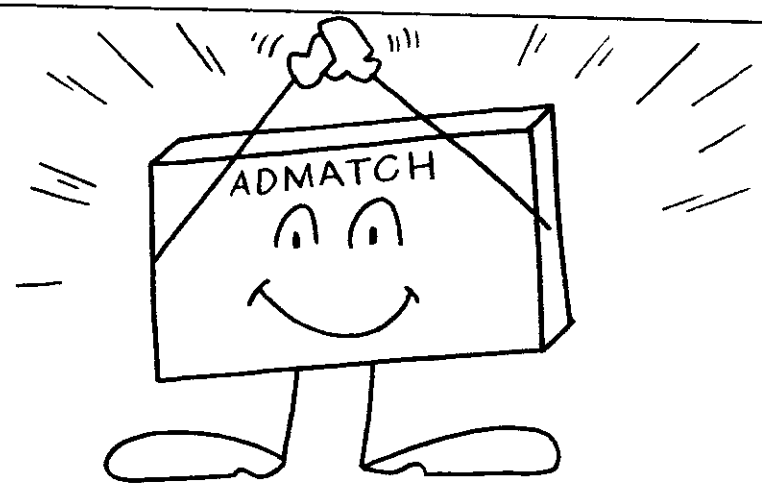
GREEN ST. — GREENE ST.
EVERETT ST. — EVRIT ST.
EIGHTH ST. — 8TH ST.

FURTHER COMPLICATIONS:

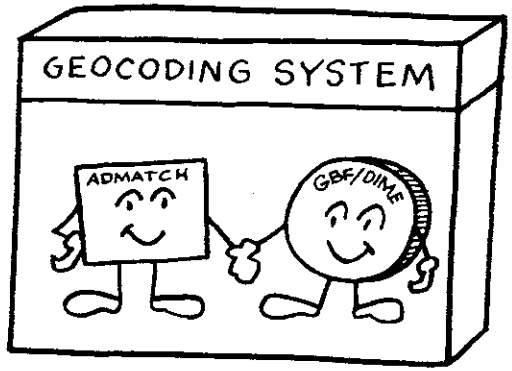
④ MISSING ADDRESS COMPONENTS:
101 MAIN
101 MAIN ST } ?

⑤ DIFFERENT RECORDING CONVENTIONS:
0025 SOUTH ST
25 SOUTH ST

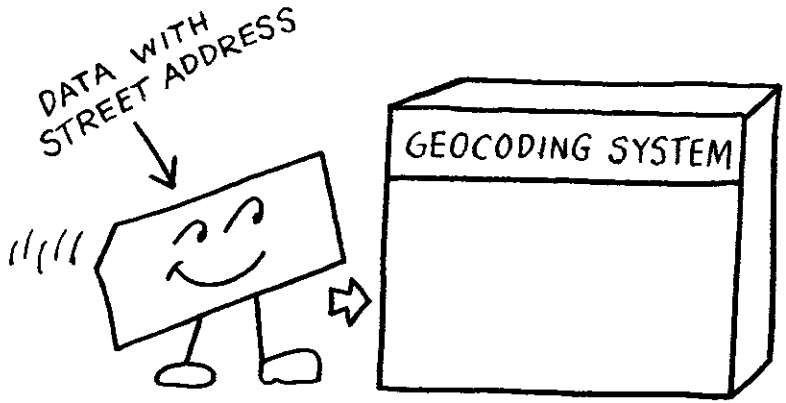
ADMATCH IS A COMPUTER SYSTEM WRITTEN EXPRESSLY TO SOLVE THE PROBLEMS OF ADDRESS MATCHING.



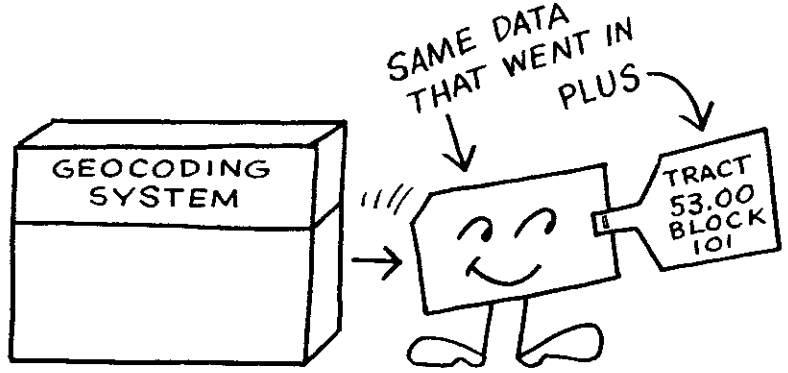
ADMATCH AND GBF-DIME WORK TOGETHER NATURALLY TO FORM A GEOCODING SYSTEM.



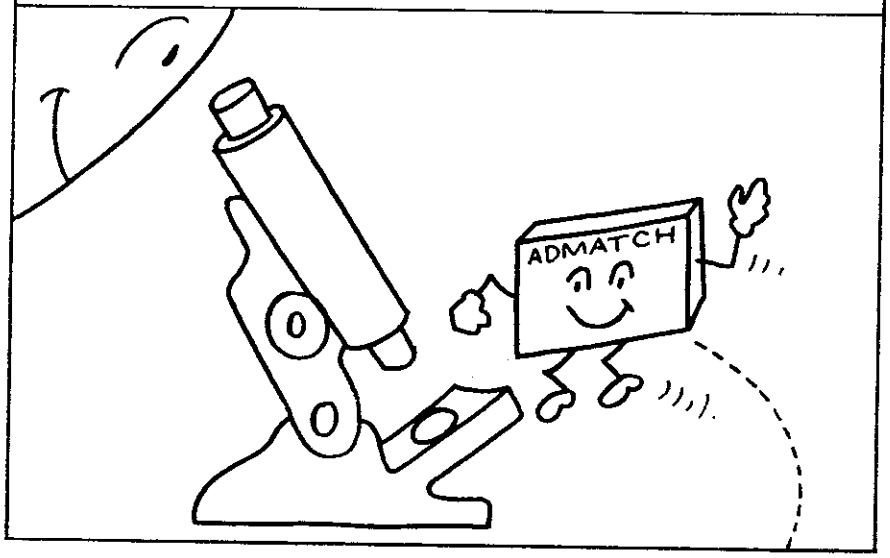
A GEOCODING SYSTEM IS DEFINED BY WHAT GOES IN.....



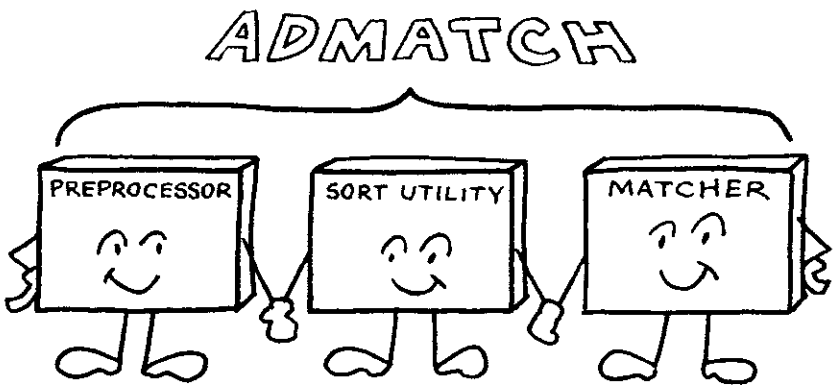
AND WHAT COMES OUT....(INPUTS AND OUTPUTS)



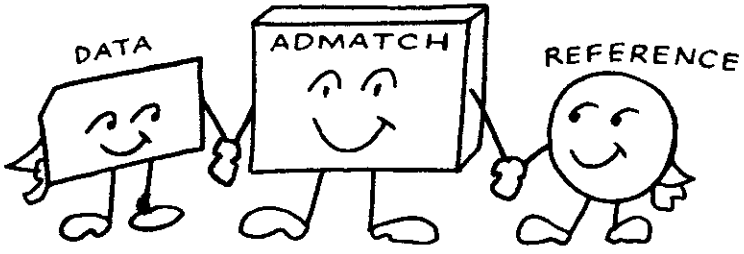
LET'S GET A CLOSER LOOK AT HOW ADMATCH WORKS.....



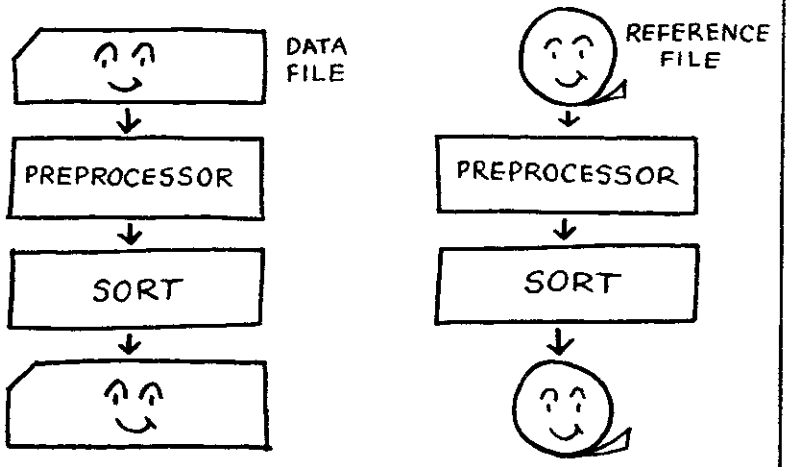
THE ADMATCH SYSTEM CONSISTS OF THREE PROGRAMS:



ADMATCH'S JOB IS TO MATCH A DATA FILE AND A REFERENCE FILE.



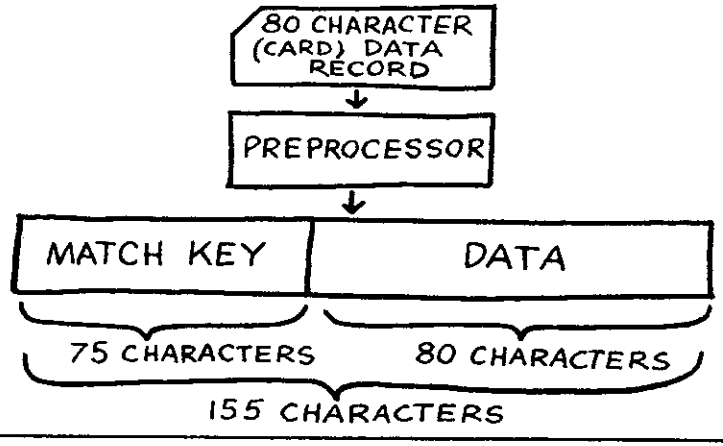
BEFORE THE ACTUAL MATCHING TAKES PLACE, BOTH THE DATA AND REFERENCE FILES MUST BE PREPROCESSED AND SORTED.

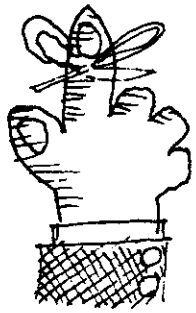


THE PREPROCESSOR'S FUNCTION IS TO STANDARDIZE THE ADDRESS OR ADDRESS RANGE ON THE DATA OR REFERENCE FILE.

BEFORE PREPROCESSING	AFTER PREPROCESSING
27 MAIN ST	00027 MAIN ST
45 MAIN STR	00045 MAIN ST
0128 MAIN STREET	00128 MAIN ST
46 WELESLEY AV	00046 WELLESLEY AV
200 WELLESLEY AVE	00200 WELLESLEY AV
179 WELLSLEY AVENUE	00179 WELLESLEY AV

THE PREPROCESSOR DOESN'T CHANGE THE ADDRESS ON THE DATA RECORD; IT COPIES THE RECORD WITH A PREFIXED "MATCHED KEY" CONTAINING THE STANDARDIZED ADDRESS.



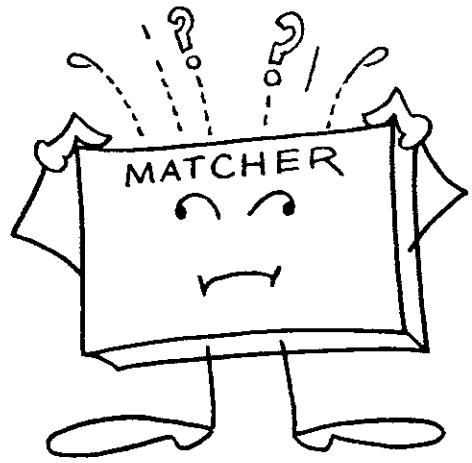


REMEMBER, BOTH THE DATA AND REFERENCE FILES MUST BE PREPROCESSED. BOTH END UP WITH A 75 CHARACTER-LONG PREFIX MATCH KEY. THE SAME STANDARDIZING RULES APPLY TO DATA AND REFERENCE FILES TO MINIMIZE DIFFERENCES.

THE NEXT STEP, SORTING, IS DONE TO MAKE THE MATCHING JOB MORE EFFICIENT.

IF THE FILES AREN'T SORTED THE MATCHER HAS A TOUGH JOB. . . . .

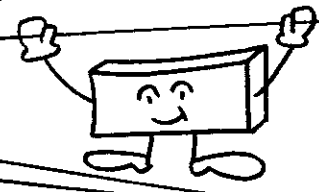
DATA	
53	ELM ST
2	CEDAR ST
88	MAIN ST
75	ELM ST
120	SOUTH ST
17	CEDAR ST
102	CEDAR ST
177	PINE ST
45	WALNUT ST
64	MAIN ST
45	MAIN ST



REFERENCE	
1-99	CEDAR ST
200-298	MAIN ST
2-98	MAIN ST
1-99	ELM ST
101-199	CEDAR ST
201-199	WALNUT ST
101-199	PINE ST
2-98	SPRUCE ST
100-198	CEDAR ST
1-99	CEDAR ST
2-98	ELM ST

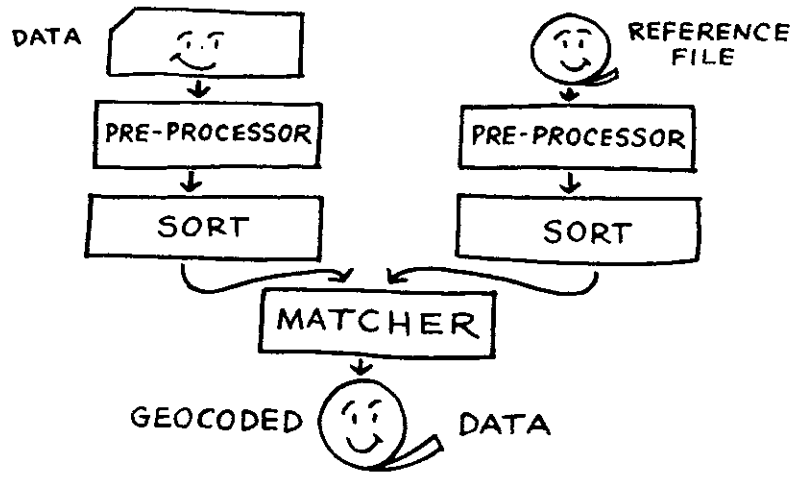
MATCHING IS EASIER WITH BOTH FILES SEQUENCED. THIS IS AN ADMATCH REQUIREMENT.

DATA	
2	CEDAR ST
16	CEDAR ST
102	CEDAR ST
53	ELM ST
75	ELM ST
164	MAIN ST
188	MAIN ST
45	MAIN ST
177	PINE CT
120	SOUTH AV
45	WALNUT ST

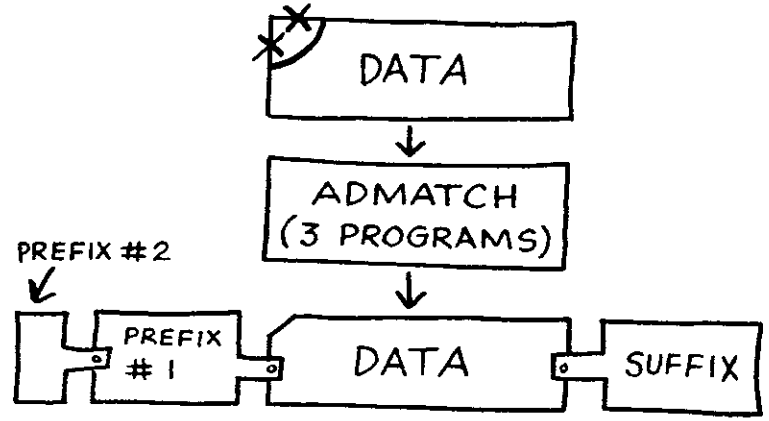


REFERENCE	
2-98	CEDAR ST
100-198	CEDAR ST
1-99	CEDAR ST
101-199	CEDAR ST
2-98	ELM ST
100-198	ELM ST
1-99	ELM ST
101-199	ELM ST
100-198	MAIN ST
200-298	MAIN ST
1-99	MAIN ST

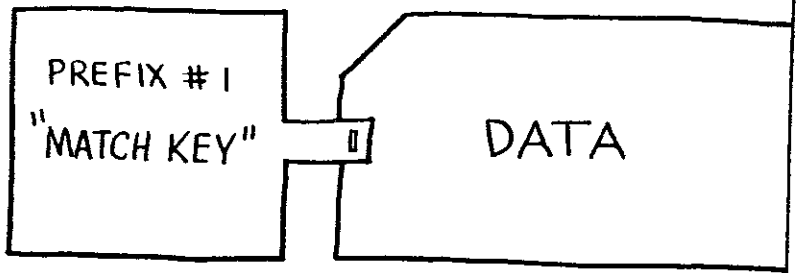
AFTER THE PREPROCESSING AND SORTING STEPS, THE MATCHER FINALLY DOES THE ACTUAL GEOCODING.



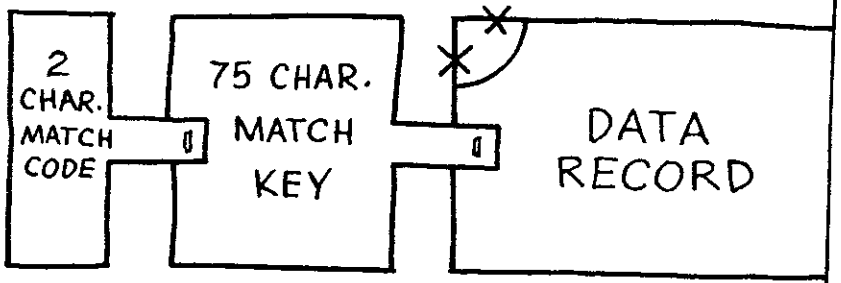
THE DATA RECORDS GO THROUGH THE ENTIRE PROCESS UNCHANGED, ALTHOUGH PREFIXES AND SUFFIXES ARE ATTACHED.



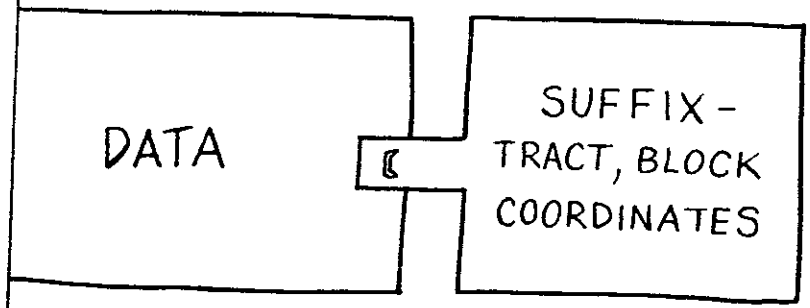
PREFIX NO. 1 ON THE FINAL OUTPUT IS THE 75-CHARACTER LONG MATCH KEY ADDED BY THE PREPROCESSOR. IT CONTAINS THE STANDARDIZED ADDRESS.



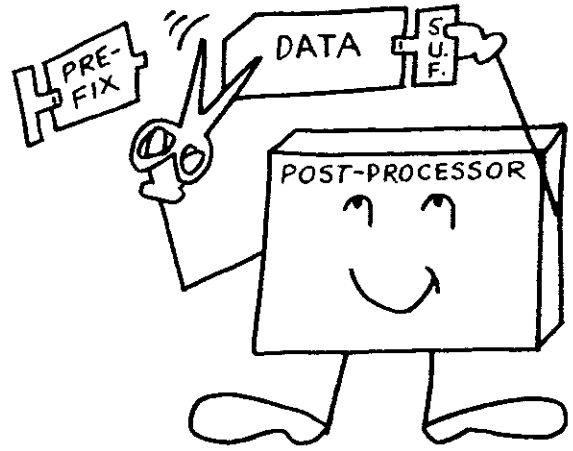
PREFIX NO. 2, IN FRONT OF PREFIX NO. 1, IS A 2-CHARACTER CODE, INDICATING WHETHER THE RECORD WAS MATCHED SUCCESSFULLY, AND IF NOT, THE REASON WHY.



BEHIND THE DATA IS THE SUFFIX, WHICH CONTAINS THE GEOCODES FROM THE MATCHED RECORD IN THE REFERENCE FILE.



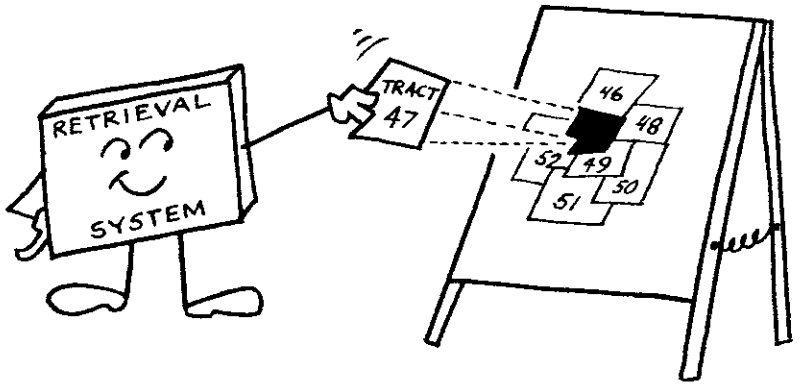
THERE IS EVEN A "POST-PROCESSOR" TO REMOVE THE PREFIXES AND/OR REFORMAT THE MATCHED RECORD.



THE MAIN THING THAT ADMATCH DOES IS TO TAKE DATA WITH STREET ADDRESSES, LOOK UP THE CORRECT GEOCODES IN A REFERENCE FILE, THEN WRITE OUT THE DATA, UNCHANGED BUT FOR THE ADDITIONS OF THE GEOCODES.

SO WHAT CAN BE DONE WITH THE ADMATCHED DATA?

SINCE ALL DATA RECORDS HAVE TRACT AND BLOCK, WE CAN RETRIEVE AND DISPLAY ALL RECORDS WITHIN A TRACT, A BLOCK, OR A GROUP OF TRACTS AND BLOCKS.



IF WE CAN SELECT DATA FOR TRACTS, WE CAN GENERATE COUNTS (OR TALLIES OR TABLES) BY TRACT.

HOUSING VALUE BY TRACT
DATA FROM GEOCODED ASSESSOR FILE

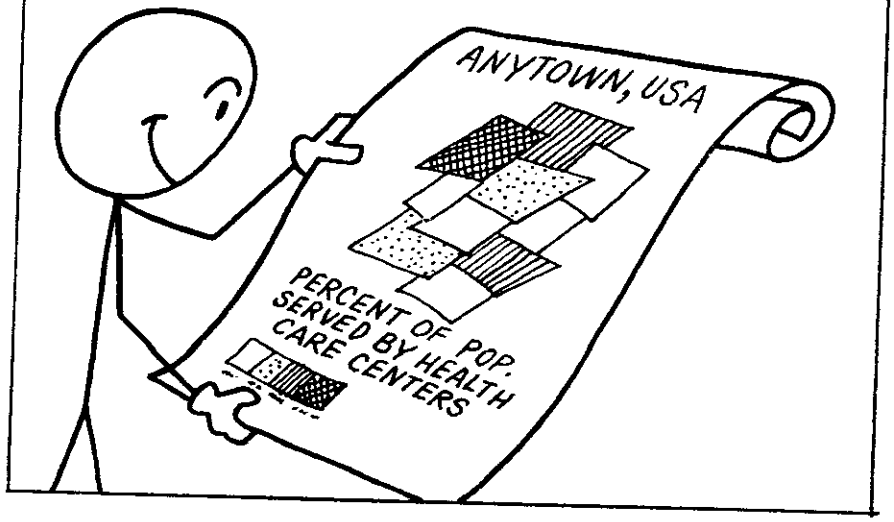
HOUSING VALUE (IN THOUSANDS)

	0-5	5-10	10-15	15-20
TRACT 5001.00	47	22	7	11
TRACT 5002.00	20	14	0	4
TRACT 5003.00	17	26	18	8
TRACT	22	17	0	12

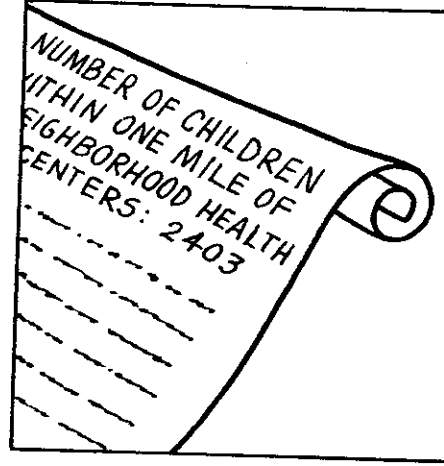
WE CAN USE 1970 CENSUS DATA BY TRACT AS "DENOMINATOR" DATA FOR PERCENTAGES— COMBINING THE CENSUS AND LOCAL DATA AND ENHANCING THE VALUE OF BOTH.

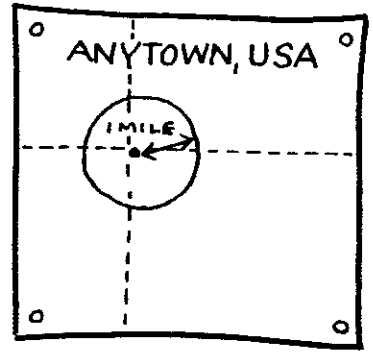
CENSUS TRACT	1970 POPULATION	HEALTH CARE RECIPIENTS	% OF POP. SERVED BY CENTERS
5001.00	3721	44	1.2
5002.00	4408	21	.48
5003.00	2650	57	2.1
5004.00	4963	122	2.5
5004.01	3910	17	.44

THE TRACT TALLIES PROVIDE INPUT FOR CHOROPLETH (FLAT-TONE) MAPPING PROGRAMS.

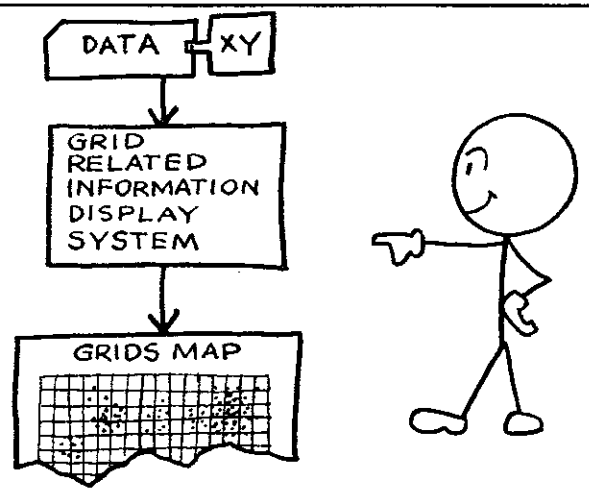


IN ADDITION TO TRACT AND BLOCK, THE COORDINATES PICKED UP BY THE ADMATCH PROCESS CAN BE USED TO SELECT DATA WITHIN A CERTAIN DISTANCE OF A POINT . . . .

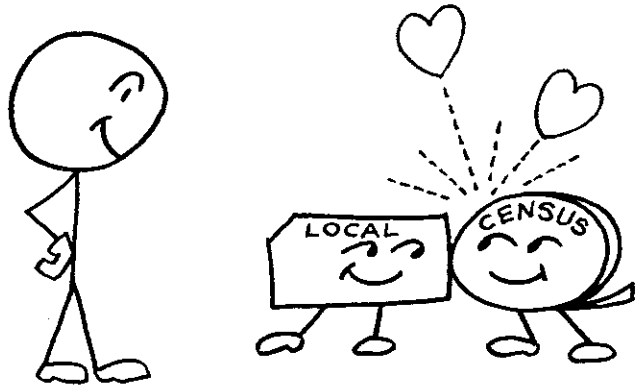




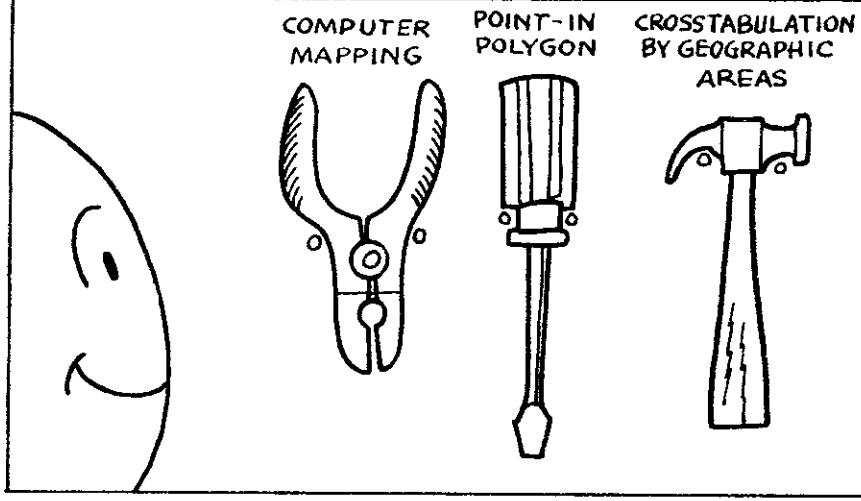
THE COORDINATE-CODED DATA GOES DIRECTLY INTO GRIDS, FOR TALLY TO GRID CELLS AND DISPLAY.



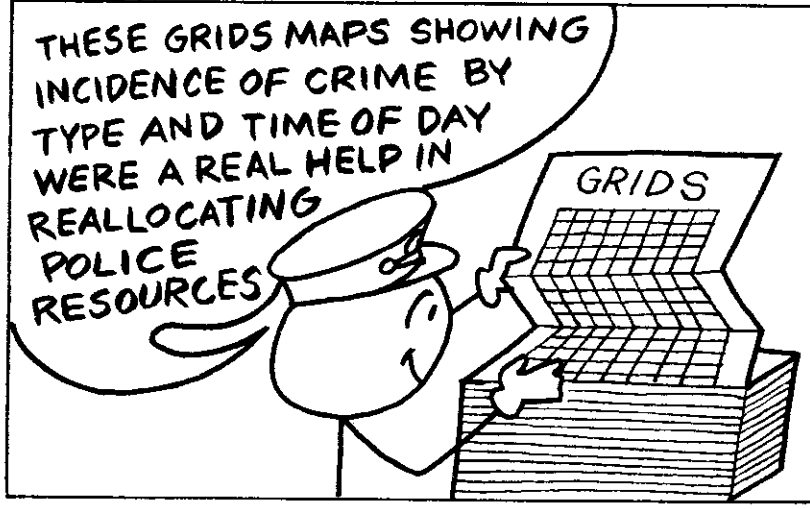
IN FACT, WHEN LOCAL DATA ARE ADMATCHED, THE GEOGRAPHIC RESTRICTIONS ARE REMOVED. THE ANALYST IS NO LONGER LIMITED BY DISPARATE GEOGRAPHY.

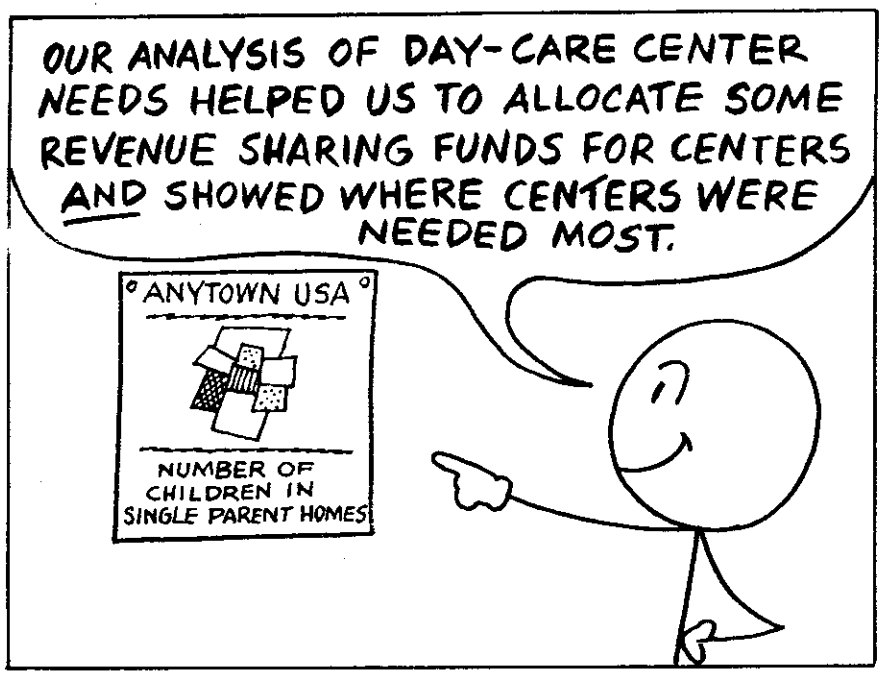


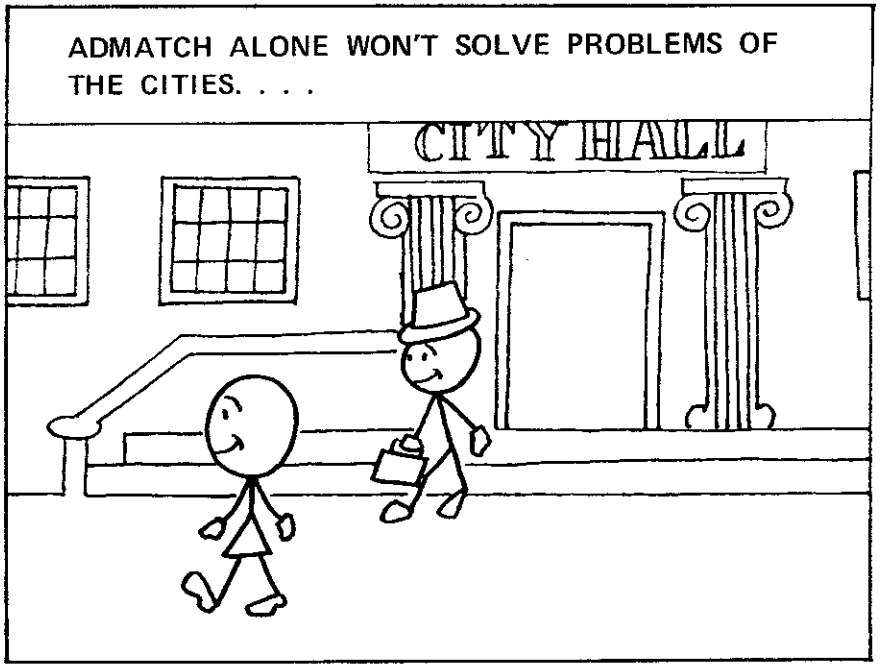
THE ANALYST CAN RELATE ALL DATA TO MANY DIFFERENT LEVELS OF GEOGRAPHY, AND THEREFORE USE AVAILABLE ANALYTICAL TOOLS.

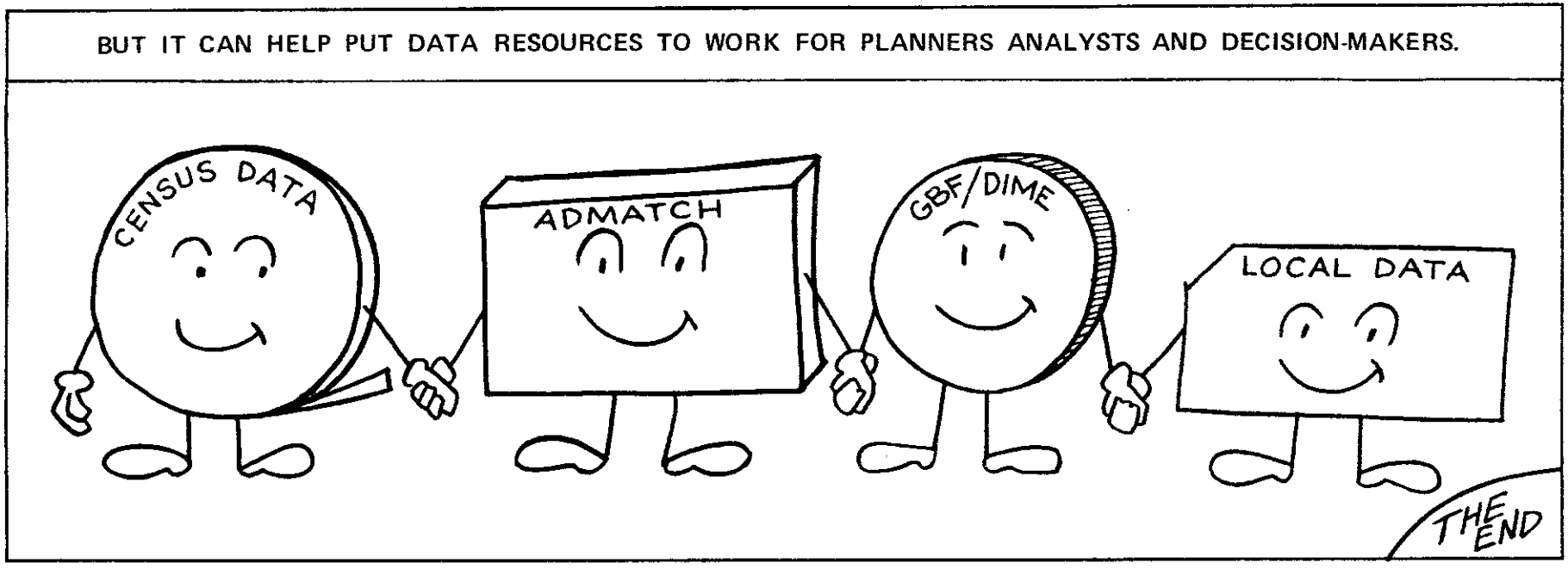


AND THE CITY EMPLOYEES HAVE USEFUL INPUTS TO THEIR DECISION-MAKING AND REPORTING PROCESSES.




OUR ANALYSIS OF DAY-CARE CENTER NEEDS HELPED US TO ALLOCATE SOME REVENUE SHARING FUNDS FOR CENTERS AND SHOWED WHERE CENTERS WERE NEEDED MOST.
ANYTOWN USA
NUMBER OF CHILDREN IN SINGLE PARENT HOMES


ADMATCH ALONE WON'T SOLVE PROBLEMS OF THE CITIES. . . .
CITY HALL


BUT IT CAN HELP PUT DATA RESOURCES TO WORK FOR PLANNERS ANALYSTS AND DECISION-MAKERS.
CENSUS DATA
ADMATCH
GBF/DIME
LOCAL DATA
THE END

GPO 874-757