COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

D.T.E. NO. 01-20

REQUEST: Verizon Massachusetts Information Requests to AT&T Communications of New England, Inc.

DATE: Original Response: May 29, 2001
Supplemental Response: September 21, 2001

VZ-ATT 1-20: State exactly what portions of HAI 5.2a's customer location database have been pre-processed and what portions are developed through running the model itself.

Respondent: R. Mercer

RESPONSE: The customer location database is produced by PNR.

SUPPLEMENTAL RESPONSE: The customer location database is preprocessed entirely by TNS (formerly PNR). No portion of the customer location database is developed by HM 5.2a-MA through running the Model.

A detailed description of the customer location process is provided in Section 5.3 of the HM 5.2a-MA Model Description. Portions of Section 5.3 describing the customer location process are repeated below:

5.3 Customer Counts by Census Block and Wire Center

Customer locations must be associated with CBs as well as their serving wire center. The PNR National Access Line Model, Version 2.0 ("NALM") performs both of these tasks. The PNR NALM uses PNR survey information, Telcordia Technologies' LERG, Business Location Research ("BLR") wire center boundaries, Dun & Bradstreet's ("D&B") business database, Metromail's household database, Claritas' 1996 demographic database, and U.S. Census estimates to calculate both the number of residential and business locations and access lines in each CB, and in each wire center in the United States. This summary describes the methodology, data and assumptions used in developing these location and line estimates in the NALM.

5.3.1 Residence Counts

Residential customer location counts are developed by applying the following process:

a) The Metromail household database (described in section 5.4.1, below) is geocoded to the "point" level.[1] In addition to recording the precise six-decimal place latitude and longitude of this household, the CB associated with its location is recorded as well. Duplicate household information is identified and eliminated. If two records appear with an identical latitude, longitude and phone number, one of the two records is eliminated.

b) Implied residential household counts are evaluated by comparing Metromail counts to Claritas' 1996 CBG-level projections of households with telephones. When Metromail households exceed Claritas households, Metromail households are used. When Claritas households exceed Metromail households, Claritas households are used, and the total differences are distributed to the constituent CBs in proportion to 1990 U.S. Census household distributions.

c) Access line counts are determined from household counts using probabilities: that is, how likely is it that a household will have a first or second telephone line installed? First line probabilities are provided by Claritas based on demographic age and income profiles by CBG. Second line probabilities are based on a logistic regression using similar demographic information and developed by PNR using its ReQuest? III residential survey. Multiplicative probability factors are applied to the household counts defined above to derive residential line counts.

d) The above derived residential line counts by CB are then normalized to sum to Study Area wide data on total residential line counts developed in Section 5.2, above.[2]

---

[1] As described in more detail in Section 5.4.3, below, geocoding to the "point" level means that the geocoding software has both found the housing unit's address in its location files and determined a latitude and longitude for the location down to six decimal places of a degree.

[2] If comprehensive LEC data on residential line counts by individual wire center are available, normalization can be done at the individual wire center level.

e) This line normalization factor is applied to the residential customer location counts in each CB, as well.

The implications of the foregoing process are as follows. Because the primary source of residential location counts is Metromail – which includes all residences that receive direct mail regardless of whether they have telephones or not – the universe of "populated" CBs that the data process captures may include CBs where telephone service is not currently offered or accepted. Thus, the "breadth" of the telephone network that these data will instruct the HM 5.2a-MA to construct is likely greater than the embedded networks of the ILECs. However, because the counts of lines and locations in each CB are normalized to sum to given Study Area wide totals, the "depth" of the constructed network will be consistent with current levels of actual telephone demand.[3]

5.3.2 Business Counts

Business location counts are developed by applying the following process:

a) The D&B national business database (described in Section 5.4.2, below) is geocoded to the "point" level. In addition to recording the precise six-decimal place latitude and longitude of businesses, the CB associated with each business's location is recorded as well.

b) From the D&B national database, the total number of business lines, as well the probabilities of these lines being single line business lines and multi-line business lines such as Centrex and PBX lines are developed. This model is based on an 800,000 firm sample.

c) Because the D&B national business database contains records for only about 11 million of an estimated total of 12 million U.S. businesses, and because the businesses that it misses are almost certainly small businesses, an additional one million non-D&B business locations are added to CB counts in proportion to D&B businesses located in the CB. The lines associated with these

[3] In addition, note that these primary source residential location counts derive from precisely geocoded 1997 Metromail data. Thus, these data provide 1997 information on location counts at the CB level. As a result, the model's reliance on noncurrent or nonCB-specific location data (e.g., Claritas 1996 CBG-level projections, 1990 U.S. Census CB counts, or 1995 U.S. Census Update county-level projections) is limited to those locations that show up in such counts that are in excess of the Metromail counts.

added business locations are projected by PNR based on an assumption that they employ, on average, between 1 and 4 employees, each.[4]

d) The above derived business line counts by CB are then normalized to sum to Study Area wide data on total business line counts developed in section 2.1, above.[5]

5.3.3 Location and Line Counts by Wire Center

HM 5.2a-MA uses wire center boundaries provided by BLR as its primary source to define wire center service areas. These boundaries conform to CB boundaries, and customer locations contained within all of the CBs associated with a wire center are then assumed to be served by that wire center.

The customer location approach introduced in HM 5.0 and continued in HM 5.2a-MA is fundamentally different from that of versions of the HAI Model prior to HM 5.0 – or any other approach that uses arbitrary geographic delineators such as CBs, CBGs or latitude and longitude grid cells. Because HM 5.2a-MA's approach identifies the actual locations of a large majority of telephone customers to within fifty feet from the center of the roads on which they are located, it produces the most sophisticated demographic data set of its type. The process first develops a database of about 109 million customer address records. These addresses are then geocoded (assigned latitude and longitude coordinates). These locations are then divided among wire center serving areas based on geocoded customer location and the BLR wire center boundaries.

5.3.4 Residence Location Data

Data for residence locations are provided by Metromail, Inc. The Metromail National Consumer Database© ("NCDB") is a large, nationally compiled file of U.S. household-level consumer information that includes deliverable postal addresses (and telephone numbers, when available). The file is compiled primarily from telephone white pages directory data, but also uses other primary sources of information, such as household mover records, voter registration data, motor vehicle registration

[4] To the extent that the D&B database contains firms that are not locatable to the CB-level, these firms are assumed to be distributed across CBs in proportion to located firms within D&B. Their line counts are calculated based on the company characteristics (e.g., employees, SIC) that they report to D&B.

[5] Again, if comprehensive LEC data on business line counts by individual wire center are available, normalization can be done at the individual wire center level.

information, mail-order respondent records, realty data, and home sales and mortgage transaction information. The file consists of close to 100 million records – which constitute over 90% of all residential housing locations that the U.S. Bureau of the Census reported for 1995.[6]

To ensure that the data captured are the most accurate available, the file undergoes numerous "hygiene" measures to ensure its continued high quality for direct marketing purposes. Such purposes require the data to reflect postal address standardization practices, incorporate National Change of Address ("NCOA") processing, and permit postal geocoding to street address, ZIP+4 or Carrier Route levels.

5.3.5 Business Location Data

Dun & Bradstreet collects information on more than 11 million business establishments nationwide. Information is gathered from numerous sources such as business principals, public records, industry trade tapes, associations, directories, government records, news sources, trade organizations, and financial institutions.

The information is organized by D-U-N-S number, a nine digit identification sequence that allows for the placement of companies within larger business entities according to corporate structures and financial relationships. A D&B family tree may be used to relate separate operating companies to each other, and to their ultimate parent company. D&B also provides "demographic" information on each of the firms in its database. Such information includes counts of employees and the Standard Industry Classification ("SIC") code of the establishment.

5.3.6 Geocoding

Geocoding is used in order to most accurately assign known customer locations to physical locations. Geocoding is also known as location coding. It involves the assignment of latitude and longitude coordinates to street addresses. Geocoding software is sophisticated enough to provide information regarding the source and precision of the latitude/longitude coordinates selected. This precision indicator allows PNR to select only those addresses that have been geocoded to a highly precise point location. Geocoding also allows customer location points to be assigned in a less granular fashion to the CBG level, or higher. Almost uniformly,

[6] This number is also very close to the 101 million households that the FCC finds in 1996, and exceeds the 95 million households that the FCC reports had telephones in that year. See, *Trends in Telephone Service*, FCC Common Carrier Bureau, Industry Analysis Division, March 1997, Table 1.

geographical address locations are derived from enhanced versions of the U.S. Census Bureau's Topologically Integrated Geographic Encoding and Referencing ("TIGER") database.

To perform its geocoding, PNR uses a program by Qualitative Marketing Software called Centrus? Desktop, which allows geocoding on two levels. The first is a match to the actual address—which is the only type of geocoding used in HM 5.2a-MA customer location. The second is a match to a ZIP code (ZIP, ZIP+4, ZIP+2) level. Because of the lesser accuracy in the second method, these geocodes are not used in PNR's process of assigning customer locations.

Within the geocode process, there are a number of options available to the user. Each of these options determines the quality of the matches allowed in the end-use geocode. For purposes of customer location, addresses are always matched to the "Close" setting. "Close" allows for minor misspellings in addition to incorrect or missing directionals (North, East, etc.) or street types (street, road, etc.). Although ZIP-based geocodes are generally accurate enough for most applications, they are not considered good enough for actual customer locations and are not used to develop and locate customer clusters in HM 5.2a-MA

Data hierarchy in address geocoding starts with the State. The hierarchy continues with City, Street Name, Street Block, and finally, House Range. Typically, a Street Block is the same as an actual physical block but it can also represent a partial block as well. The House Range displays address information from the USPS. Additionally, where there are gaps in the actual address range, the House Range will account for these gaps.

Initially, the address coding module in Centrus Desktop compares the street addresses from the input file to the records contained in the USPS ZIP+4 directory and the enhanced street network files. If the address is located in the USPS files, the address is standardized and a ZIP+4 is also returned. If this address is also found in the street network files, Centrus Desktop determines a latitude and longitude for the location. Optionally, if the address is not found in the street network files, location information may be applied from the ZIP level.[7]

Location codes generated by Centrus Desktop indicate the accuracy of the geocode. For purposes of customer location clustering in the HM 5.2a-

[7] Note that ZIP+4 codes may be very precise. In general, they are specific to the face of single city block. While it may turn out that accuracy to the street block face is quite sufficient for accurate cost modeling of local telephone networks, in the interest of conservatism, these type of geocodes are not presently used in HM 5.2a data.

MA only those geocodes assigned at the 6-decimal place point location made directly to the street segment are used.[8]

While the software and data used allow for a much more comprehensive output of data elements, for use in HM 5.2a-MA customer location, the following addressing elements are extracted:

Address
City
State
ZIP
ZIP+4
Latitude
Longitude
Census Block
Match Code
Location Code

5.3.7 Gross-up

The above-derived precisely geocoded locations are counted by CB. These geocoded location counts by CB are then compared to target total line counts for that CB derived by the PNR NALM (as described in Section 2.3, above). If the geocoded location counts are less than the target count, the residual number of customer location points is then computed, and geographical locations for these points are generated. This process is performed by PNR using TIGER file road locations. Each of the additional number of customer location points that a CB requires to total to its target count is generated and assigned a geocode so as to place these "surrogate" points uniformly along roads within the CB,[9] giving double weight to interior roads. Placing surrogate locations in this fashion provides a conservatively high dispersion of customer locations because it does not account for further clustering within a CB, such as small villages or neighborhoods separated by green space.

[8] Furthermore, placement of the address along the street segment is quite precise. The Centrus geocoding software and reference data also make use of USPS determinations of whether the segment contains a continuous or discontinuous range of address numbers. Thus, if the addresses on a block face run from 200 to 250 and 274 to 298 (with the range between 252 and 272 missing), an address of 250 will be geocoded, it will not simply be geocoded as at midblock.

[9] Road locations are taken from the TIGER files, excluding certain types of roads along which customers are unlikely to be located, such as: limited access highway segments, road segments that are in tunnels or underpasses, vehicular trails and roads passable only be four-wheel-drive vehicles; highway access ramps, ferry crossings, pedestrian walkways and stairways, alleys for service vehicles, and driveways.

As a result of this gross-up process, the customer location file now contains records for each of the U.S.'s more than 100 million customer locations with a geocode (either calculated precisely or through the gross up process) associated with it.

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

D.T.E. NO. 01-20

REQUEST: Verizon Massachusetts Information Requests to AT&T Communications of New England, Inc.

DATE: Original Response: May 29, 2001
Supplemental Response: September 21, 2001

VZ-ATT 1-21: For the State of Massachusetts provide:

a. the number of addresses obtained through the Metromail, Inc. National Consumer Database;

b. the percentage of addresses to total households obtained through the Metromail, Inc., National Consumer Database; and,

c. the percentage of addresses that are P.O. Boxes and Rural Route Boxes.

Respondent: R. Mercer

RESPONSE:

a. The requested information is based on copyrighted Metromail, Inc., data, and is commercially available from Metromail.

b. See response to part "a".

c. The Model does not report percentage of addresses that are P.O. Boxes or Rural Route Boxes.

SUPPLEMENTAL RESPONSE: Without waiving the objection set forth in its initial response, AT&T states as follows:

The response to this information request contains proprietary information and is being provided only to the Department and parties which have signed a confidentiality agreement with AT&T.

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

D.T.E. NO. 01-20

REQUEST: Verizon Massachusetts Information Requests to AT&T Communications of New England, Inc.

DATE: Original Response: May 29, 2001
Supplemental Response: September 21, 2001

VZ-ATT 1-24: Describe in detail and provide all documents concerning, referring or relating to the exact procedure used by HAI 5.2a to normalize line counts by census block to sum the Study Area wide data on total residential line counts as described at pages 25-26 of the Model Description. This description should state the basis for additions or reductions to specific census blocks that are made in order to perform the normalization of total line counts for the study area to the targets.

Respondent: R. Mercer

RESPONSE: The procedure for normalizing line counts is described in detail in Section 5 of the HAI 5.2a-MA Model Description.

SUPPLEMENTAL RESPONSE: As Section 5.3.1 of the HM 5.2a-MA Model Description makes clear, residential line counts by Census Block developed as described in steps (a) through (c) are summed to produce an area-wide total. This total is normalized to the Study Area wide data on total residential line counts developed as described in Section 5.1[1] of the HM 5.2a-MA Model Description – namely, for Verizon, data taken from the 1999 ARMIS 43-08 report. This leads to a normalization factor that could be less than or greater than unity. For instance, if the sum of the residential line counts by CB were only 90% of the study area wide residential line count

[1] The HM 5.2a Model Description provided as Exhibit RAM-2 of Dr. Mercer's Direct Testimony contains a typographical error in step (d) of Section 5.3.1 in that it refers to study area residence line counts developed as described in "Section 5.2," rather than the correct reference to "Section 5.1." Similarly, step (d) of Section 5.3.2 refers to study area business line counts developed as described in "Section 2.1," rather than the correct reference to "Section 5.1."

reported in ARMIS data, the normalization factor would be 1/.9 = 1.11. This normalization factor would then be applied to the line counts in each CB, causing them to increase by 1.11. Similarly, if the normalization factor were less than unity, the line count in every CB would be scaled down by this factor. The same process occurs for business lines, as described in Section 5.3.2 of the Model Description.

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

D.T.E. NO. 01-20

REQUEST:	Verizon Massachusetts Information Requests to AT&T Communications of New England, Inc.
DATE:	Original Response: May 29, 2001 Supplemental Response: September 21, 2001
VZ-ATT 1-25:	Provide all documents concerning, referring or relating to the estimated total business count of 12 million that is used as the basis for the business adjustment referenced at page 27 of the Model Description.

Respondent: R. Mercer

RESPONSE:	The requested information is based on copyrighted Dun and Bradstreet data. The information is commercially available from Dun & Bradstreet.
SUPPLEMENTAL RESPONSE:	No documentation exists which supports the development of an estimated total business count exceeding the approximate 11 million businesses recorded in the Dun and Bradstreet national business database.

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

D.T.E. NO. 01-20

REQUEST: Verizon Massachusetts Information Requests to AT&T Communications of New England, Inc.

DATE: Original Response: May 29, 2001
Supplemental Response: September 21, 2001

VZ-ATT 1-26: Provide all software and inputs that constitute the PNR Associates, Inc. ("PNR") clustering algorithm.

Respondent: R. Mercer

RESPONSE: See response to VZ-ATT 1-23.

SUPPLEMENTAL RESPONSE: Without waiving the objection set forth in its initial response, AT&T states as follows:

The clustering algorithm software of TNS Telecoms (formerly PNR Associates, Inc.) is attached. One copy of the CD-ROM is being provided to the hearing officer and one copy is being provided to Verizon. If any other party would like a copy of the CD-ROM, AT&T will provide one. The attachment contains proprietary information, however, and will only be provided to parties which have signed a confidentiality agreement with AT&T.

Should Verizon wish to run the software using the TNS geocoded data set for Massachusetts, AT&T repeats its offer to make the necessary arrangements with TNS to provide Verizon with (1) remote electronic access to the geocoded data set; and (2) the necessary support and documentation in order to facilitate the use of the attached software.

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

D.T.E. NO. 01-20

REQUEST: Verizon Massachusetts Information Requests to AT&T Communications of New England, Inc.

DATE: Original Response: May 29, 2001
Supplemental Response: September 21, 2001

VZ-ATT 1-78: Provide any and all documents concerning, referring or relating to the engineering, furnishing, and installation of AT&T's most recently constructed power plant including the addition of rectifiers, batteries, fuse distribution bays, automatic breakers, microprocessor, and the standby emergency generator.

Respondent: R. Mercer

RESPONSE: See objection in response to VZ-ATT 1-70.

SUPPLEMENTAL RESPONSE: See AT&T's supplemental response to DTE-ATT 1-4, which will be filed shortly.

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

D.T.E. NO. 01-20

REQUEST: Verizon Massachusetts Information Requests to AT&T Communications of New England, Inc.

DATE: Original Response: May 29, 2001
Supplemental Response: September 21, 2001

VZ-ATT 1-79: Identify any and all expenses concerning, referring or relating to the installation of AT&T's most recently constructed power plants, including riggers, transportation, and heavy equipment as well as all installation labor costs.

Respondent: R. Mercer

RESPONSE: See objection in response to VZ-ATT 1-70.

SUPPLEMENTAL RESPONSE: See AT&T's supplemental response to DTE-ATT 1-4, which will be filed shortly.

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

D.T.E. NO. 01-20

REQUEST: Verizon Massachusetts Information Requests to AT&T Communications of New England, Inc.

DATE: Original Response: May 29, 2001
Supplemental Response: September 21, 2001

VZ-ATT 1-82: Explain in detail how HAI 5.2 "locates" customers who are not identified through the geocoding process. Provide any and all documents concerning, referring or relating to the process of locating such customers.

Respondent: R. Mercer

RESPONSE: The requested explanation is provided in Section 5.3.7 of the HAI 5.2a-MA Model Description and pages 39-40 of Dr. Mercer's testimony.

To the extent that the question is seeking any software or documentation that is the intellectual property of PNR, see response to VZ-ATT 1-23.

SUPPLEMENTAL RESPONSE: Without waiving the objection set forth in its initial response, AT&T states as follows: Please see the attached *Ex Parte* Presentations to the Federal Communications Commission:

1. *Ex Parte Submission by AT&T Communications, Inc., and MCI Telecommunications Corp.*, CC Docket Nos. 96-45 and 97-160, September 30, 1997; and

2. *Ex Parte Presentation by AT&T Communications, Inc.*, CC Docket No. 96-45, March 2, 1998 (includes a description of alternative "road" surrogating methodology).

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

D.T.E. NO. 01-20

REQUEST: Verizon Massachusetts Information Requests to AT&T Communications of New England, Inc.

DATE: Original Response: May 29, 2001
Supplemental Response: September 21, 2001

VZ-ATT 1-83: Provide the "clustering algorithm" that is used to determine groupings of customers and explain how this algorithm is utilized in HAI 5.2a and all previous versions of the Hatfield Model Release. Provide all documents concerning, supporting, referring or relating to this algorithm.

Respondent: R. Mercer

RESPONSE: The clustering algorithm process is described in Section 5.4.2 of the HAI 5.2a-MA Model Description.

To the extent that the question is seeking any software or documentation that is the intellectual property of PNR, see response to VZ-ATT 1-23.

SUPPLEMENTAL RESPONSE: Without waiving the objection set forth in its initial response, AT&T states as follows: Please see the Supplemental Response to VZ-ATT 1-26 and the attached *Ex Parte* Presentations to the Federal Communications Commission:

1. *Ex Parte Presentation by AT&T Communications, Inc.*, CC Docket No. 96-45, January 13, 1998 (includes a copy of the C++ code for PNR's Spatial Clustering Module);

2. *Ex Parte Presentation by AT&T Communications and MCI Telecommunications Corp.*, CC Docket No. 96-45, January 13, 1998 (includes a "description of how the Hatfield Model, v. 5.0, performs at counting, locating and clustering customers, and at engineering the local telecommunications network");

3. *Ex Parte Presentation by AT&T Communications and MCI Telecommunications Corp.*, CC Docket No. 96-45, January 30, 1998 (includes "several pages of the C++ code of the Spatial Clustering Module…that highlight some of the key mathematics of the clustering algorithm"); and

4. *Ex Parte Submission by AT&T Communications, Inc., and MCI Telecommunications Corp.*, CC Docket Nos. 96-45 and 97-160, September 30, 1997 (provides the source code and a licensing agreement for the clustering algorithm) (attached to Supplemental Response to VZ-ATT 1-82).