THE LRMS CROSS STREETS PROFILE WITH COORDINATES

TECHNICAL EVALUATION

United States Department of Transportation
Federal Highways Administration
Office of Safety and Traffic Operations
ITS Research Division
Contract DTFH61-91-Y-30066

Vehicle Intelligence & Transportation Analysis Laboratory
University of California at Santa Barbara
September 1998

CAUTION !!

This is a draft document. Typographic errors are known to exist in some tables of results.

The Cross Streets Profile with Coordinates

—

Technical Evaluation

FINAL REPORT

United States Department of Transportation
Federal Highways Administration Contract DTFH61-91-Y-30066

Project executed under contract to
Viggen Corporation, Interop Division
Knoxville TN

With infrastructure funded by
California Department of Transportation
Testbed Center for Interoperability

Val Noronha
Peter Fohl
Michael Goodchild
Richard Church

Vehicle Intelligence & Transportation Analysis Laboratory
National Center for Geographic Information and Analysis
University of California
Santa Barbara CA 93106-4060

September 1998

Table of Contents

This report describes tests of the LRMS Cross Streets Profile, carried out by the Vehicle Intelligence and Transportation Analysis Laboratory (VITAL) at the University of California, Santa Barbara. Tests were performed in two phases. The original Cross Streets Profile (Goodwin et al 1996) was tested first. Following our preliminary recommendations, the profile was strengthened with coordinates (SAE 1998). A second phase of testing was then required, to examine the enhanced Profile.

The Profile does not specify implementation details, hence there is no single/correct approach to testing. Our experiments took a number of approaches to illuminate various aspects of performance. Phase One tests were carried out with little error-checking and strict enforcement of rules. These produced success rates in the 5–15% range. In Phase Two, with more sophisticated algorithms, and leniency on rules, results in the 50–75% range (on average) were achieved.

Whether or not these results are acceptable depends upon the application. Clearly mission-critical tasks such as emergency servicing require a higher success rate. Success achieved by relaxing the rules may be misleading, therefore it is important that the Profile incorporate a measure of messaging reliability. When coordinates are used in conjunction with street names, the message contains two independent forms of location expression. One interpretation of the coordinate component is that it ensures “success,” because even if there is a 200m positional error, it establishes at least an approximate location. Another interpretation is that corroboration between coordinate and street name components constitute a test of reliability, in the absence of which the message is suspect.

Our approach to testing Phase Two focuses on success as defined and established by corroboration. Two complementary experiments form the core of this phase. First, a position is relayed purely via coordinates, and cross street names are used only to judge the validity of the transfer. Second, cross street names are processed in increasingly sophisticated ways, with assistance from coordinates, and success is judged by (i) Euclidean distance measures (i.e. using coordinates) and (ii) the frequency with which the name processing fails and the process reverts to coordinates. Samples are stratified to deal separately with major streets and urban areas.

The first series of tests — transfer by coordinate alone — produces a “probable” success rate of 11% over the County of Santa Barbara, with a further 61% of transfers rated “possibly” accurate. Using a limited set of points sampled in the field around the *City* of Santa Barbara, the corresponding numbers are 24% and 26% respectively. Note that this series is not a realistic model of real world implementation, because street names are used merely to evaluate, not to assist the process. In the second series of tests, which uses both names and coordinates, about 50–85% of name matching attempts are successful, the rate on major streets being *lower* at 10–60%. Most of the transfers (40–96% on all streets; 50–97% on major streets) are within 30m of the source point — typically indicating an accurate transfer.

Clearly there are far too many dimensions to the tests to be able to quote a single success statistic. The maximum scores in the 90s are impressive, whereas the lower scores in the 10% and 50% range are a matter of concern. Our test effort took a variety of approaches, to explore as many aspects of the problem as possible. One has to interpret the results according to one’s needs.

At first sight the Cross Streets Profile appears to be a logical method for communicating a location, and it is surprising that success rates should be as low as they are. The principal reason for message failure is database quality:

- Between 20% and 45% of all records in current commercial databases for Santa Barbara County have blank name fields. This is particularly problematic for the Cross Streets Profile which relies heavily on street names.
- Some streets that are in one database are not in another. In particular, freeway ramps may not be faithfully represented, traffic circles (“roundabouts”) may be generalized into regular 4-way intersections, right-turn slip lanes may not be shown, driveways into residential complexes and institutional areas may or may not be included.
- There are no standards for structuring street names. One database may label a street North Main Street, others may cite it as Main Street, or Main Street North. Vendors structure the components of street name differently, and even established standards such as US Postal Service street type abbreviations are not universally respected. Software can be designed to cope with some of these problems; standardization between vendors would be a better solution.

We anticipate that database quality will improve over time, and that commercial service providers will devote considerable effort to develop more advanced algorithms than we did under the constraints of the study. In the interim, other strategies such as the ITS Datum may have to be pursued.

Glossary

LRMS	Location Referencing Message Specification (Goodwin et al 1996)
VITAL	Vehicle Intelligence & Transportation Analysis Laboratory, University of California, Santa Barbara
XSP	Cross Streets Profile as originally proposed in LRMS
XSP2	Revised Cross Streets Profile including coordinates (SAE 1998)

In August 1997 the Vehicle Intelligence Testing & Analysis Laboratory (VITAL) was contracted by Viggen Corporation, Tennessee, on behalf of the United States Department of Transportation, Federal Highways Administration, to test the Cross-Streets Profile (XSP). The XSP is part of the Location Reference Messaging Specification (LRMS) proposed by Viggen and Oak Ridge National Laboratories (ORNL) for ITS messaging (Goodwin et al 1996). VITAL's original report (VITAL 1997b) documented significant problems in the XSP, and recommended changes, one of which was to include coordinates in the XSP. The profile was subsequently modified (SAE 1998), and VITAL performed a new series of tests.

This document covers both analyses. Phase One is the test of the original Profile; Phase Two describes additional experiments following the enhancements with coordinates. This document is designed to be read in conjunction with Viggen's consolidated LRMS test report (Haas et al 1998). It therefore assumes that the reader is familiar with ITS principles and terminology, and the background to the LRMS and messaging standardization efforts.

Persons primarily responsible for the design and execution of the tests, and generation of this report, are Val Noronha and Peter Fohl. Principal investigators on VITAL projects are Michael Goodchild and Richard Church. The research was undertaken in cooperation with Viggen, ORNL and the California Department of Transportation (Caltrans). We are grateful for discussions with Ramez Gerges, Cecil Goodwin, Steve Gordon, Robert Haas and John Lau.

Further details on the project are available from VITAL at

NCGIA/Department of Geography
University of California
Santa Barbara, CA 93106-4060

or from the VITAL web site, http://www.ncgia.ucsb.edu/vital

Contact persons are

1. Dr. Val Noronha
   Phone +1.805.893.8992
   E-mail noronha@ncgia.ucsb.edu

2. Professor Michael Goodchild
   Phone +1.805.893.8049
   E-mail good@ncgia.ucsb.edu

3. Professor Richard Church
   Phone +1.805.893.4217
   E-mail church@geog.ucsb.edu

Infrastructure to support this work was developed under an earlier contract funded by the Caltrans Testbed Center for Interoperability.

Introduction

A process for communicating a location from one map system to another must be accurate, unambiguous and efficient. There are several ways to describe a location, for example:

- Coordinates (e.g. latitude, longitude)
- Route and distance (e.g. Interstate 95, mile 253)
- Intersections (State Street and Hope Avenue)
- Landmarks (Cottage Hospital)
- Map references (page 35, E-6)

Each method has its advantages and disadvantages, which make it appropriate to a particular set of applications.

The Cross Streets Profile

The LRMS specification proposes standardization of message content and format. It describes a number of options or messaging "profiles." The Cross Streets Profile (XSP) (Goodwin et al 1996) specifies a point location in terms of

a) the street on which the point lies (Birch St in Figure 1),
b) the intersecting streets (Ash and Cedar) that bound the segment on which the point lies,
c) the offset distance (absolute or normalized) of the point from either of the intersections in (b).

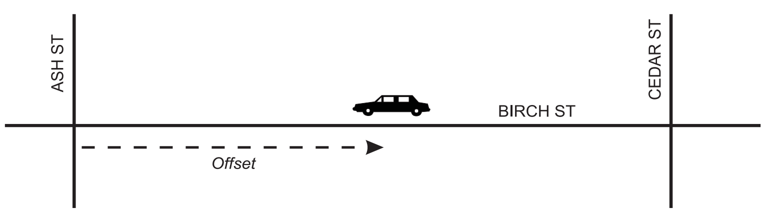

Figure 1. The Cross Streets Profile

Conceptually, the XSP describes the location of the vehicle in Figure 1 as follows: "875 metres [or a normalized distance: 55% of the way] along Birch St between Ash St and Cedar St." In LRMS nomenclature, Birch is the *On-Street*, Ash and Cedar are the *From* and *To* streets respectively.

The XSP provides for identification of a point on a street, or the entire street segment.

Test Strategy

The system receiving the XSP message must be able

a) to find the street and both cross-streets unambiguously, and
b) to determine the distance offset to a satisfactory level of accuracy.

Evaluation of (a) is straightforward: either the intersections are correctly identified, or they are not. This can be measured by a "hit rate," defined as the proportion of successes in identifying the correct intersection.

A measure of effectiveness of (b) is not as clear, for a couple of reasons. First, if the location is a large landmark such as a museum or a public beach, a motorist can easily identify it visually, and 200–500m may be sufficient accuracy. On the other hand, the point may be a curve speed advisory sign on a highway, with a tolerance is 2–5 metres. Secondly, at the receiving end the point may be stored as a 2-dimensional (x,y) coordinate. If there are significant disagreements in the geometry of the street between databases, the (x,y) location is distorted, and subsequent transmission of this distorted position to another database could result in gross misrepresentation of the original data. Since the ultimate effects of error therefore cannot be determined in terms of "successful" versus "unsuccessful," a scientific test can only report on the likelihood and severity of error. The implications of this must then be interpreted within the context of a given application.

Accordingly the XSP can be viewed as having two components: a topological component that identifies the street intersection which serves as a reference location, and the coordinate geometry component that resolves the position of the point along the street. These components have to be tested separately, because the issues surrounding them are entirely independent of each other. Clearly both (a) and (b) must fall within acceptable ranges. No matter how good the positional accuracy of the data bases [i.e. good score on (b)], a poor hit rate on (a) renders the XSP method ineffective.

Test Data Bases

Vendors

Our tests are based on commercial databases for the County of Santa Barbara, California. Data were acquired directly from vendors in May 1997 under license. All databases were ordered in ArcInfo® export format. Layers were filtered as required to exclude features not directly pertaining to the street network, such as landmarks, hydrography, etc. Coordinates were standardized to NAD83.

We sought six databases for testing. They are identified by letter symbols A–F. Database C was an engineering database which, although positionally accurate, lacked the topological structure required for XSP use. Database D was not available at the time of testing.

The databases are intended to illustrate in the most practical way the types and severity of error that would be expected if the XSP were deployed today. VITAL is not in the business of ranking vendors, or conducting comparative studies of the databases.

Differences between Databases

When a XSP message is received, the first task for the receiver is to search the destination database to find the triad of names that define the street segment (in Figure 1 these are Birch St, Ash St and Cedar St). In theory this should be a simple text matching operation, a technique widely discussed in the computer science literature. Unfortunately, due to missing or blank data fields, and differences in vendor practice, the name matching problem takes center stage in the XSP.

First, about a third of the street records, accounting for 40–60% of the total length of roads, have missing or blank names (Table 1). A position cannot be transmitted in XSP unless a valid name exists for *all three* relevant streets. This means that about a third of all potential location transfers are disqualified. The XSP could fail on this point alone.

Table 1. Incidence of blank street names

	Vendor A	Vendor B	Vendor E	Vendor F
Records	35483	32687	20067	30000
Non-Blank	19965	22050	16223	17661
% Blank	44	33	19	41
Distinctive	3913	3493	4136	3105
% Blank by length	60	47	43	58

Second, vendors represent and parse street names in different ways. A name could potentially be broken down into at least five fields: Directional prefix, Street Type prefix, Proper Name, Street Type suffix, Directional suffix. Table 2 lists examples of how these components are applied (local variations exist, that may not be accommodated within this structure):

Table 2. Street name components

Dir Prefix	Type Prefix	Proper Name	Type Suffix	Dir Suffix
		Main	Street	
		Main	Street	NW
North		Main	Street	
	Interstate	95		
	Avenue of the	Americas		
	Via del	Cuadro		

Of the four vendors studied, only one uses such a comprehensive format. Others use a variety of formats as indicated in Tables 3–5.

Table 3: Vendor A Fields

Proper Name
N ARRIBA WY
AVENIDA REDONDO
N BLOSSER RD
PINE CANYON RD

Table 4. Vendor B/F Fields

Dir Prefix	Proper Name	Type	Dir Suffix
N	ARRIBA	WAY	
	AVENIDA REDONDO		
N	BLOSSER RD		
	PINE CANYON	RD	

Only Vendor A stores an alias field. Typically this is used where a highway number (US-101) coexists with a text name (El Camino Real); in some cases, but not always, it is used for an alternate spelling (Lynne Road vs Lynn Road), or to strip off a directional prefix (N Arriba vs Arriba).

Not all vendors respect U.S. Postal Service street type abbreviation standards — one uses Arriba Way, another uses Arriba Wy. For a search algorithm to equate Way with Wy (and similarly Ave with Av), street types must be parsed out and compared separately from other name subfields.

However, due to the non-standard separation of fields as illustrated in the tables, parsing is complex, the rules being dependent on local practice.

Table 5: Vendor E Fields

Dir Prefix	Type Prefix	Proper Name	Type Suffix	Dir Suffix
N		ARRIBA	WY	
	AVENIDA	REDONDO		
N		BLOSSER	RD	
		PINE CANYON	RD	

Commercial software exists for large scale name matching of streets and addresses. Our mandate in testing the XSP is not to re-invent such software. We break down the name matching problem into levels of sophistication as follows:

Level 1

- Concatenate name components as appropriate, compare case insensitive:
  AVENIDA REDONDO = AVENIDA REDONDO

While Level 1 should ideally include examination of aliases, only Vendor A provides for an alias field.

Level 2

- Concatenate/parse name fields as appropriate
- Standardize and equate street types (Av = Ave)
- Expand abbreviations (N = North, St = Saint)
- Ignore embedded spaces and punctuations (Cottonwood = Cotton Wood)

Level 3

- Collapse ordinal words to numeric ordinals (5th = Fifth)
- Test for containment (Ward Memorial = Clarence Ward Memorial)

The above levels resolve differences in *vendor practice* and operator-level interpretation of a street name. Higher level processing (e.g. phonetic matching and transposition of words/characters) can be implemented to resolve discrepancies that fall within the realm of *errors* e.g. Birch vs Birck. As higher levels of probabilistic matching are introduced, there is a chance that spurious matches will occur. For example, Market St equates with Marquette St using Soundex phonetic matching. Since the XSP requires three names to score a match on a street segment, the likelihood of spurious matches is diminished.

XSP Problem Scenarios

Recall the conceptual XSP message of Figure 1: “875 metres [or a normalized distance: 55% of the way] along Birch St between Main St and State St.” Consider some situations in which this would result in an ambiguous or incorrect position at the receiving end.

1. Due to errors or differences in vendor practice, “Birch St” cannot be found in the receiving data base.

    - The receiving database uses an alias for Birch St: Hwy 43
    - Spelling error: Birck St
    - Vendor practice:
        - conflict in abbreviation or coding — Av vs Ave; Route 43 vs Hwy 43
        - prefix/suffix problems
    - Vendor interpretation: Ward Memorial Blvd vs Clarence Ward Memorial Blvd
    - Human error: Birch St coded as Birch Av

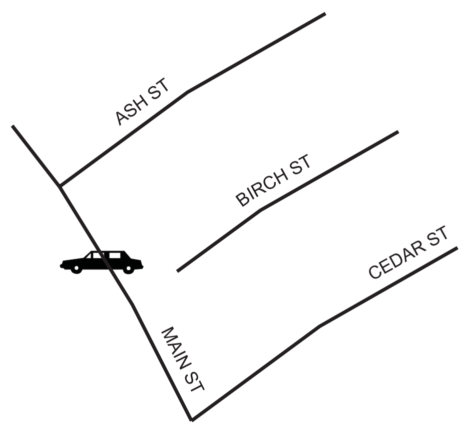

Figure 2. Topological error situation

2. Topological Mismatch. In Figure 2, Birch St comes within a few metres of Main St, but does not intersect it. Suppose the transmitting database erroneously records an intersection — a common error — whereas the receiving database does not. Then a message that uses this intersection cannot be interpreted at the receiving end. In the converse situation, where an intersection exists in the receiving data base but not the transmitting system, intelligent receiver software can infer the location correctly, although with most link-oriented data structures in current use, this is difficult.

3. A more general form of (2) occurs when a street is missing entirely from the source or target database. While errors of commission/omission are found in all kinds of neighborhoods, new residential areas and intricate freeway interchanges are prime candidates.

4. An absolute or relative offset measured along the on-street inevitably translates to slightly different positions in the transmitting and receiving data bases. The extent of discrepancy depends on the resolution and accuracy of the data bases. The tolerance to error depends on the application.

The above situations are due to database error or incompatibility. There are also problems inherent in the XSP itself. Figure 3 shows four departures from the normal T- or 4-way intersection configuration. In the case of two streets that intersect more than once (Figure 3a), the XSP provides for an “occurrence number” to distinguish one intersection from another; however, it does not specify how occurrences are counted and ordered. It does not explicitly cover situations (b), (c) or (d).

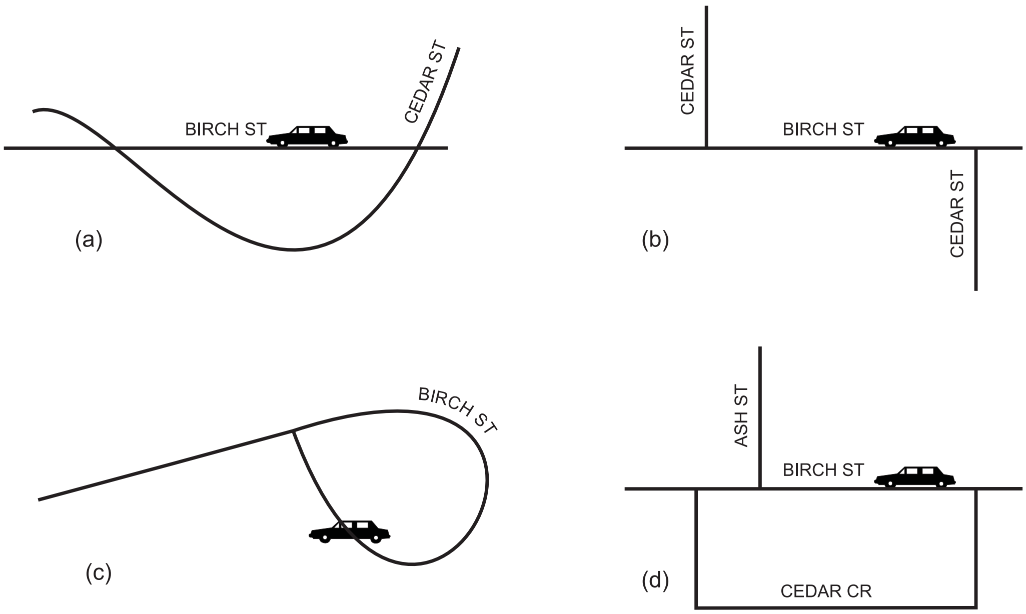

Figure 3: Multiple-Intersection Ambiguities

Definitions and Evaluation Metrics

The issue to be tested is the effectiveness with which location information can be transmitted using the XSP. The evaluation focuses not so much on message field layout and format, as on the ability of the XSP to perform in the context of municipal naming practices, and the errors typically found in commercial databases. The following constitute unsatisfactory performance:

- the message cannot be interpreted by the recipient (e.g. a reference to a street name that does not exist in the recipient's data base), or
- the interpreted location lies beyond a tolerance distance of the original point, or
- the message is interpreted incorrectly, resulting in an incorrect location.

(1) is measured by a success- or hit-rate, (2) in terms of Euclidean displacement. (3) usually results in high positional error, and is trapped in tests for (2). It is difficult to automate tests for the few instances of relatively limited significance, in which (3) results in positional error that falls within tolerance.

Intersection: For the purpose of the XSP test, an intersection is defined as a point at which two or more streets with differing names intersect. Bi-valent intersections (usually pseudo-nodes) are ignored, except where the two incident streets have different names — this often occurs at elbows.

Hit Rate is the proportion of transmissions in which the intended result is correctly interpreted at the receiving end. The hit rate applies only to the topological component of the XSP, i.e. identifying the correct segment of the on-street; it does not apply to the offset. A hit is not scored if the message is ambiguous (multiple hits). If the recipient finds exactly one intersection, other than the one intended to be transmitted, this is more serious, in that the parties may be unaware of the error — this is considered a failure. Since it is impractical to verify the validity of each hit, a failure is recorded if the distance between the intersection points in the source and target databases is greater than a critical distance, say 300m. There may be instances of failures (e.g. when two streets intersect more than once over a short distance) that are not detected by this criterion. Hit rate statistics are explained in greater detail in the following section.

Accuracy is a measure of positional difference between the point transmitted and that received. The significance of this difference is context-dependent. The XSP provides only for points that lie *on* the network (*near*-network points such as gas stations and hotels are generalized and snapped to the network), therefore the spatial frame of reference is abstracted. Motorists operate in this abstracted space, with respect to street names and intersections, not coordinates. Therefore it could be argued that it is sufficient for navigational purposes for the motorist to receive only a relative position along a street, e.g. Hollister, 200 metres west of Patterson. As long as this position is resolvable in the receiving data base, coordinate accuracy is irrelevant.

However, broader principles of interoperability require that coordinate-based systems (e.g. GPS) communicate with an offset-based system. Therefore it is necessary to preserve spatial (coordinate) accuracy for its own sake, and it is appropriate to measure error in terms of positional displacement.

Accuracy is presented as follows:

- For absolute offsets:
  - difference in normalized offset of the point in the source and target databases;
  - difference in cartesian coordinates: magnitude of displacement vector.
- For normalized offsets:
  - difference in absolute offset, and
  - difference in cartesian coordinates: magnitude of displacement vector.

Experimental Procedure and Analytical Results

VITAL developed software to implement the XSP. Two types of tests were devised: field observations and lab simulations. In general, lab simulation is an inexpensive way to generate a large number of test points. However, any simulated points have to be generated with respect to, or snapped to, the source street network prior to transmission, and are therefore arguably biased towards the source database. With field observations there is no bias since the points are independently surveyed using differential GPS, but there are practical limits on the number of test points that can be surveyed.

Field Survey Design

VITAL's test infrastructure is described in VITAL (1997a). Using our field data gathering facility we surveyed about 60 points within about a 50 km radius of the city of Santa Barbara, the area of survey being limited only by the reception range of differential GPS signals. Points were sampled along highways and city streets, at recognizable objects such as specific entrances to buildings such as motels, or traffic signs. The test vehicle was photographed at each point, showing the GPS antenna in relation to the fixed object, in the event that the points would need to be revisited in the future.

At each point, the following data were recorded

- On-street name
- From-street name
- To-street name
- GPS reading

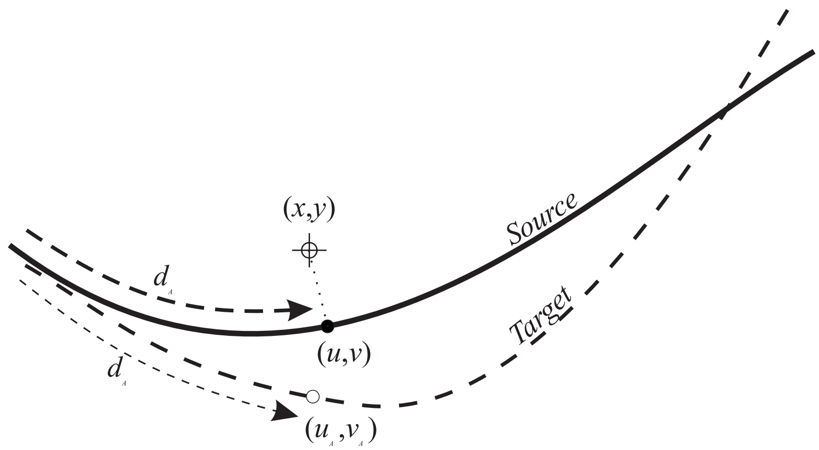

Figure 4. Processing of Field Test Measurements

Databases were studied in pairs, one database being a source and the other a target. For each pair (A→B, A→C, etc):

1. Each GPS reading (x,y) was snapped to the nearest centerline segment in the source database to derive (u,v). Offsets d_A (absolute) and d_N (normalized) were calculated from the intersection with the from-street.
2. The snapped point (u,v) was relayed by means of the XSP, using absolute and normalized offsets, to the target database.
3. If the triad of street names could not be found in the target database, the transfer was abandoned. If the target street was found, the target database coordinates were inferred from the offsets: (u_A, v_A) absolute and (u_N, v_N) normalized. The normalized offset with respect to the target database d'_N was calculated from (u_A, v_A), and absolute offset d'_A was calculated from (u_N, v_N).
4. Displacement vectors were measured:
   - Absolute offset-derived coordinates $(u-u_A, v-v_A)$
   - Normalized offset-derived coordinates $(u-u_N, v-v_N)$
   - Offset error: absolute $d_A-d'_A$
   - Offset error: normalized $d_N-d'_N$
5. Coordinate displacements greater than 300m were considered to have originated from spurious name matches, i.e. failures.

Lab Test Design

The simulated test points were processed in much the same way as the field surveyed points. Clearly a much larger sample, about 10,000 points, was possible. Test points were generated at 1 km intervals along streets in each source database. This sampling strategy ensured that the density of test points reflected network density. It also meant that a different set of test points was generated for each source. While short streets obviously had fewer points selected along them, streets of equal length had an equal likelihood of having points selected along them.

1. Since the test points were generated from the centerline, they each lay on a centerline segment, subject only to rounding error. These points were equivalent to (u,v) above. Offsets d_A (absolute) and d_N (normalized) were calculated from the intersection with the from-street.
2. The test point (u,v) was relayed by means of the XSP, using absolute and normalized offsets, to the target database.

3. If the triad of street names could not be found in the target database, the transfer was abandoned. If the target street was found, the target database coordinates were inferred from the offsets: (u_A, v_A) absolute and (u_N, v_N) normalized. The normalized offset with respect to the target database d'_N was calculated from (u_A, v_A), and absolute offset d'_A was calculated from (u_N, v_N).
4. Displacement vectors were measured:
   - Absolute offset-derived coordinates (u-u_A, v-v_A)
   - Normalized offset-derived coordinates (u-u_N, v-v_N)
   - Offset error: absolute d_A-d'_A
   - Offset error: normalized d_N-d'_N
5. Coordinate displacements greater than 300m were considered spurious name matches, i.e. failures.

Test Results

Shake-down Test

The name matching component of the XSP algorithm was initially tested exhaustively in a small study area about 7x5 km in Goleta CA. Results were examined manually to ensure validity. This test area is entirely urban, therefore the problem of blank names is not as severe as in rural localities. Yet the proportion of records that could be tested ranged from 36% to 58%.

Table 6 lists the results of name matching in Goleta. The following terminology is used:

- Records: The number of graphic entities in the source file. This number can be misleading because of the presence of pseudo-nodes.
- Attempts: To enable a transfer, the on-street must be non-blank, and at least one intersecting street must be present at each end, with a non-blank name different from that of the on-street. A looped street segment (Figure 3c) would be rejected.
- Hits: A hit is exactly one match occurring within a tolerance distance (300m), even after testing both directions on the target database. In Figure 3d, there are two instances of “Birch between Ash and Cedar” — this is an ambiguity.
- Ambiguities: Where there is more than one possible match, due to any of the situations in Figure 3 (a,b,d), an ambiguity is recorded.

Table 6: Name Matching Results for Goleta test area

Source Database		Target Database			
		A	B	E	F
A	Records	2277	2277	2277	2277
	Attempts	822	822	822	822
	Hits	771	236	431	15
	Ambiguities	51	11	22	3
	Hits/Attempts	94%	29%	52%	2%
B	Records	1802	1802	1802	1802
	Attempts	967	967	967	967
	Hits	240	873	252	33
	Ambiguities	9	94	12	6
	Hits/Attempts	25%	90%	26%	3%

Source Database		Target Database			
		A	B	E	F
E	Records	1731	1731	1731	1731
	Attempts	1007	1007	1007	1007
	Hits	441	252	933	32
	Ambiguities	17	12	74	6
	Hits/Attempts	44%	25%	93%	3%
F	Records	1717	1717	1717	1717
	Attempts	837	837	837	837
	Hits	17	33	32	786
	Ambiguities	2	6	6	51
	Hits/Attempts	2%	4%	4%	94%

Even when a street segment is matched back to the source database, ambiguities result in hit rates under 100%. These results are for Level 1 name matching. Better rates can be expected at Level 2; at Level 3 and above, there is a greater likelihood of spurious matches, and therefore a higher proportion of ambiguities.

Lab Tests

Table 7 shows name matching results for the approximately 10,000 points generated around the county. The hit percentages are in general slightly lower than for the Goleta shake-down test study area. This is to be expected because of the dominance of remote rural roads at the county scale.

Table 7: Name matching results for entire county of Santa Barbara

Source Database		Target Database			
		A	B	E	F
A	Records	35483	35483	35483	35483
	Attempts	8218	8218	8218	8218
	Hits	7813	2067	2731	1104
	Ambiguities	405	81	104	59
	Hits/Attempts	95%	25%	33%	13%
B	Records	32687	32687	32687	32687
	Attempts	10392	10392	10392	10392
	Hits	2070	9488	2170	3904
	Ambiguities	79	904	100	131
	Hits/Attempts	20%	91%	21%	38%
E	Records	20067	20067	20067	20067
	Attempts	10795	10795	10795	10795
	Hits	2750	2175	10247	740
	Ambiguities	94	97	548	40
	Hits/Attempts	25%	20%	95%	7%
F	Records	30000	30000	30000	30000
	Attempts	9034	9034	9034	9034
	Hits	1112	3908	738	8616
	Ambiguities	55	129	40	418
	Hits/Attempts	12%	43%	8%	95%

Table A.1 (Appendix A) presents numerical results of offset error analysis.

- Clearly the principal problem is in forming valid XSP triads. On average, only 20% of all test points have all three required name fields.
- The match rate on valid triads varies from 5% to 38%, with a mean of 18%, using Level 1 matching. The overall transfer success rate for all test points is 1% to 6%, with a mean of 3%.
- The average positional error is 40–50m. The worst case error is more than 1 km. Relative offsets produce a slightly lower positional error than do absolute offsets.

Field Tests

The lab tests above are supplemented by field tests, for two reasons:

- Many of the points generated in lab tests lie on private or gated roads. Results based on those tests are appropriate in worst-case application scenarios such as emergency management. On the other hand, field test points are sampled along public motorable roads, more representative of typical ITS-oriented driving.
- Lab tests are based on points sampled with respect to a given database. One could argue that those points are inherently biased towards the source database. Field sample points on the other hand are independently surveyed and therefore unbiased.

Numerical results of offset error analysis for field tests are shown in Table A.2 (Appendix A). In general the results are better than those observed in lab tests, but the success rate is still low: 25% from valid triads, and 14% overall. Table 8 summarizes the differences between lab and field test results.

Table 8: Lab and field results compared

	Lab	Field
Number of test points	10,000	54
Formed valid XSP name triads in source database	20%	56%
Unambiguous hits on target database from valid triads (Level 1 match)	18%	25%
Unambiguous hits on target database from all test points (Level 1 match)	3%	14%
Average transfer offset error using absolute offset (metres)	48	32
Average transfer offset error using relative offset (metres)	42	28
Worst case transfer offset error using absolute offset (metres)	1042	199
Worst case transfer offset error using relative offset (metres)	1047	194

Discussion of Results

Clearly the problem with the XSP is not so much with positional accuracy of offsets — generally 30 to 50 metres — as it is with failure to identify the target link correctly. Transfer success rates with the XSP are so low as to be unacceptable for most applications. The problem areas are

a) blank street names,
b) non-standard ways of parsing and storing street names, and
c) street configurations that cause topological ambiguities when trying to resolve intersections using the language of the XSP.

Problem (b) can be fixed to an extent with software; and (c) accounts for a small number of cases. The major problem is (a).

The Unnamed Streets Problem

Blank street names caused low hit rates across all tests. However, since hit rates are tied closely to the sampling of test points, they could be unrepresentative of actual use in some applications. In the lab tests for name matching and transfers above, points or links were selected systematically along *all* roads, hence remote country roads were as likely to be selected as were busy freeway segments. The sampling strategy is valid from a geo-statistical standpoint, but does not reflect traffic density or frequency of incidents, and may therefore be irrelevant to some applications.

Figures 5 and 6 show the distribution of unnamed streets (illustrations based on Vendor A) in south western Santa Barbara county, and in the Santa Barbara/Goleta urban area. It is clear from the maps that the vendor’s emphasis is on major public streets; many of the unnamed roads are privately owned or even gated. It could therefore be argued that the unnamed street problem is of little relevance to many ITS applications, although it a matter of concern in mission-critical applications such as emergency management.

One could advance the following arguments against the numerical results in the preceding section:

1. Hit rates should be studied with regard to major streets only, suppressing sample points on minor and private roads; or the results should be weighted by traffic volume. This point is expanded below.
2. The XSP could be applied more intelligently. Our testing has applied a basic interpretation of the XSP, i.e. we required that the target link have the required cross-street names at both ends. If an intermediate cross-street is missing or topologically inaccurate in either of the databases, the XSP fails although all locational clues could have in fact been found. There are a number of possible strategies to improve upon this, i.e. to find second-, third- or lower-order intersections that *do* have names. However, there are two potential problems with this approach:
    - There may be long stretches of road with no valid (non-blank) cross-streets. For example, California State Highway 1 runs for about 30 km without crossing a named road. Offset error would be large on such segments.
    - The computations required to perform multi-order searches for non-blank names may not be feasible on low-end processors envisaged for use in vehicles.
3. The XSP should be strengthened by passing coordinates for (a) the end-points of the road segment (currently identified by intersecting street names), and (b) the incident location. A pair of coordinates requires at most 64 bits, which is nominal in comparison to a 10- to 30-character street name. Moreover, a coordinate would limit the scale of error in the event of a spurious hit. However, this modifies the character of the XSP, and it may be preferable simply to use other profiles.

The Major Streets Argument

In the immediate future many ITS applications are likely to be restricted to freeways and major arteries. By the time they are ready for deployment on lower-order streets, vendors are likely to have improved the quality of their databases. Therefore it could be argued that statistics based on *all* streets are misleading, and that separate statistics should be gathered for *major* streets.

The first problem — both in deploying a limited system for major streets, and in testing the argument — is that there is no standard definition of a major street. Taxonomies of streets differ widely. Vendor A uses a broad but useful classification with just 8 feature classes such as Interstates, US Highways, etc; on the other hand Vendors B and F employ more than 40 categories. It is impossible to find an aggregation of features codes for each vendor, that would result in an identical subset of streets to be tested. This is particularly problematic for the XSP, because unless a road segment is defined by the same three streets in both the source and target databases, the position cannot be relayed.

Secondly, many of the naming problems outlined above apply equally to major streets, particularly the problem of blank names. One vendor stores highway numbers (US-101) as the principal name, whereas another uses text names in the principal field (Ventura Freeway), storing the highway number (US Hwy 101) as an alias. For the XSP to work reliably, the receiving system would have to be designed with due attention to these differences. Based on cursory examination of the databases, we do not anticipate hit rates dramatically higher than when all streets are considered.

The field tests were conducted on public streets, with a good representation of major streets. The results were more encouraging than those of the lab tests, but still generally poor. It is unclear whether restricting the study to major streets would push the performance levels of the XSP into an acceptable range. Further study is required on this matter.

Conclusions and Recommendations

This study demonstrates that the demands on a location messaging profile based on street names are considerable. Whereas positional data can be gathered and updated inexpensively from aerial photography, attribute data such as street name require laborious field work and verification, and are susceptible to error. Municipal and community practices in street network design and street naming are so variable, that it requires a robust standard to accommodate every idiosyncrasy.

We draw the following conclusions from the study:

1. The success rate in XSP transfers is so low as to be decidedly unsatisfactory as a general purpose interoperability strategy. The hit rate for non-blank names is in the range of 5–70%, but due to the high proportion of blanks, the overall hit rate is 1–40%. Due to the blank names problem we considered it pointless to pursue higher levels of name matching under the current project.

2. The results are specific to the XSP, and do not reflect on other message specifications that employ linear referencing, such as the LRMS Linear Referencing Profile. The principal problem with the XSP arises from the use of street names to identify a street segment; the Linear Referencing Profile is more flexible, allowing the use of indices or names.

3. Position coordinates should be included in the XSP. This would resolve ambiguities in some instances. However, it raises the question of whether the coordinate is the primary or secondary specification, *vis-à-vis* the on-, from- and to-street names. A pair of coordinates requires 64 bits for transfer at 10-metre precision; a triad of street names requires about 240 bits using ASCII (as specified by the XSP). Moreover, with the addition of coordinate information, the XSP starts to resemble the LRMS Geographic Coordinate Profile, which relies primarily on coordinates, with abbreviated street names passed as secondary information.

4. The XSP specification is incomplete on the matter of multiple intersections. Multiple intersections are not uncommon, and before the XSP is deployed as a standard it should be expanded to address this issue.

There is a chance that the XSP could be extended and adapted to improve hit rates and to make it more useful. Its effectiveness could conceivably be improved using the following techniques:

1. Higher level name matching, at least to Level 2
2. Testing on major streets only
3. Multi-order search for named intersections.

Should the XSP be pursued further, these would be appropriate subjects for future research.

Introduction

The success of the Cross Streets Profile (XSP) for location messaging is dependent on a number of factors:

- Inherent effectiveness of the Profile — The ability of the Profile to accommodate all street configurations. The Phase One report detailed problem situations in which certain configurations led to referential ambiguities.
- Implementation — The XSP does not specify how the message is composed at the transmitting end, or interpreted at the receiving end. Intelligent software may improve effectiveness by forgiving errors and incompatibilities, say in matching the strings "5th Av" to "Fifth Avenue," or in high-order topological search; it may be that one approach to interpreting a location message is inherently superior to other approaches.
- Database accuracy — The performance of the profile is ultimately a spatial data interoperability issue, directly dependent on the quality of databases.
- Compliance with standard practices — Map database vendors differ in how they parse street name into components; similarly they do not all follow USPS standards on street type abbreviations. Success rates are higher between databases that comply with common standards.
- Municipal practices — Errors are inevitable when municipal naming practices are poorly designed, e.g. multiple streets with the same name.
- Application context — EMS requirements are far more demanding than ITS or marketing requirements.

Clearly a time-constrained test effort must steer a course between a rigid, unintelligent interpretation of the Profile, and a sophisticated implementation driven by substantial investment and evolution over time. VITAL's previous work evaluating the Cross Streets Profile (Phase One) recommended the following:

- reinforcement of the Profile with coordinates
- processing of blank names
- fuzzy name search
- topological search
- stratification by street class

Phase Two testing of the revised Cross Streets Profile (XSP2) (SAE 1998) addresses these issues to varying degrees. We have effected a limited set of enhancements in software intelligence, to explore methods of improving success rates, and to demonstrate the dependence of performance statistics on the quality of implementation.

Test Dimensions

Coordinates

With the inclusion of coordinates, XSP2 contains two sets of referencing information — street names and coordinates — that could potentially be in disagreement and may therefore be considered independent. If they corroborate each other, it may be assumed that the interpretation of the location reference is correct.

The XSP2 does not dictate whether the coordinate or street name should take precedence in the event of conflict. Two fundamentally different approaches are possible:

- Interpret the message based on names, and reconcile problem cases using coordinates; or
- Use coordinates as a starting point, using names as tie-breakers.

Hybrid methods between these two extremes may be derived.

XSP2 provides for two coordinate triplets, specifying the start and end points of an object of interest on the road. A point event uses one coordinate triplet, whereas a linear event such as a no-passing zone requires two points. Again, this leads to potential conflict, in that the two points may translate to two different links in the destination database. Using street name it may be possible to establish that one is right and the other wrong; however, if blank names are encountered — up to 45% of records in some databases have blank names — the ambiguity is not resolvable. Moreover, the Profile explicitly permits negative offsets, allowing the two points to lie on different links.

Fuzzy Name Search

Selected fuzzy name matching strategies are implemented:

- Street name and street type are treated as independent fields wherever possible. In the case of database A, the fields are not separated. There are two ways to deal with this: (a) parse the components, or (b) test for containment of the name field of the other database within A's combined name field — this is problematic with short names such as "B Street".
- Street type is normalized to USPS standards, e.g. "Av" is always represented as "Ave." Street type prefixes ("Via del," "Avenida") are concatenated with street name.
- Highway notations are normalized based on examination of vendor data, e.g. "US Hwy 101" is equated with "Hwy 101" using a lookup table
- Name strings are collapsed to remove spaces and punctuation
- Alias name fields are considered where available.

This pre-processing can by no means be considered exhaustive. For example, subfields are not cross-checked — Avenue Redondo is not equated with Redondo Avenue; Main North is not equivalent to North Main.

Blank Names

In Phase One, test points were disallowed if one or more of the streets in the XSP triad had blank names. This resulted in rejection of 70–80% of all test points. Phase Two tests are designed to be more forgiving, allowing blank names under certain conditions. Take for example a search for Main between Ash and

Birch. If the triad {Main, Ash, *blank*} is encountered in the destination database, *and there is no other instance* of {Main, Ash, Birch}, then the *blank* is overlooked and assumed to match Birch. Further, triads that are completely blank are allowed to match identical triads in the destination — this allows a large number of "hits" that would not even have been admitted for Phase One testing (Table 9 shows the proportion of triads in each database that are completely blank). While Phase One represents one end of a spectrum of strictness, appropriate for say EMS applications, Phase Two tests are a significant step back from that hard-line position.

Database	% blank triads
A	31.06
B	19.32
C	2.30
D	9.97
E	27.72

Table 9: Incidence of Blank Triads

Topology

The Phase One report recommended, in the event that blank cross-street names were encountered in the source file, that the algorithm "fan out" to test adjacent links until non-blank names were found. More generally, this approach is useful when intervening streets occur in the destination database between the from- and to-streets, and for transfer of major street events.

Stratification

The purpose of stratifying streets is to test the hypothesis that (a) data for major streets are better, and (b) major streets are where most ITS incidents occur, therefore statistics based on all streets are biased.

Stratified tests may be run (a) limiting both source and destination databases to major streets, or (b) allowing all streets in both databases, but generating sample points only along major streets in the source database. Strategy (b) is adopted for the following reasons:

- There are significant differences in the ways vendors classify and code "major" streets (Figure 1). Strategy (a) would require that each database be processed differently. Then because of differences in network density, it would be difficult to find identical triads of on-streets and cross-streets in source and destination.
- Strategy (b) is more representative of reality. Motorists would typically carry the complete street database; however, incidents would tend to occur principally along major streets.

It is important not to expect too much of the major streets approach. About 10% of all "major" streets, however defined by each vendor, have blank names.

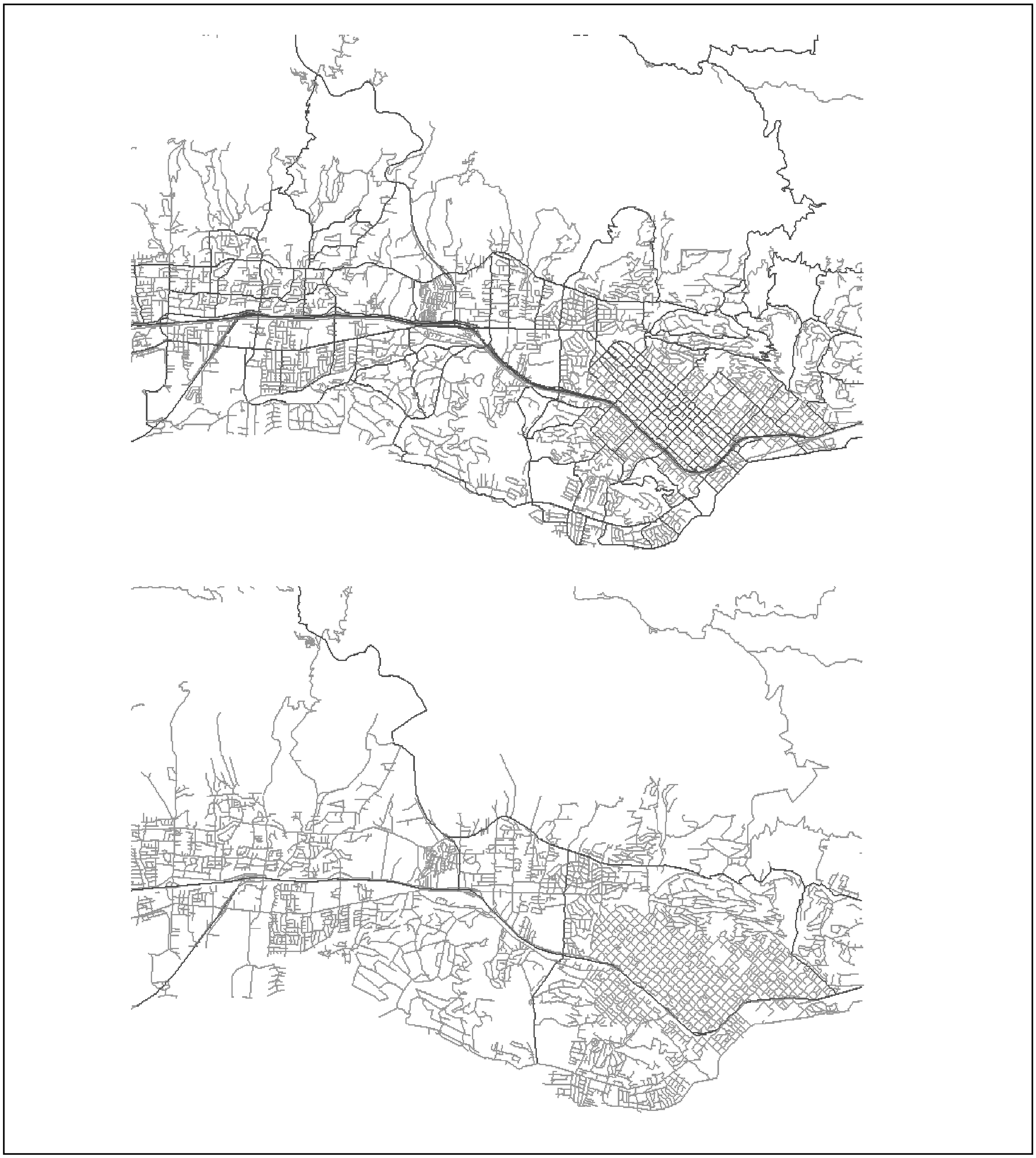

Figure 5. Urban Goleta and Santa Barbara: two vendors' versions of "major" streets

Note that freeway ramps are not uniformly classified as major, although many major incidents occur along them.

Intersection Rules

Intersection rules are clarified in Phase Two:

- The cross street at a dead-end is explicitly coded as *null* rather than *blank*.
- Pseudo-nodes destroy the integrity of a link. Since pseudo-nodes are distributed arbitrarily, they lower the success rate of XSP2. For Phase Two each data file is pre-processed to delete pseudo-nodes.

Experiment Design

A. Match to Self

The first test of the XSP is to match a location message, composed using the XSP specification, to the database from which it was derived. Two types of problems are uncovered by this test:

- Problems caused by incorrect municipal practice or database inaccuracy, such as repeated street names
- Inability of the profile to handle certain topological configurations.

These tests were performed in Phase One. Clearly, using a coordinate in conjunction with a street name in its native database, *all* messages are fully resolvable, therefore there is no need to repeat this test for XSP2.

B. Using Coordinates to Target a Link

This series of tests determines the reliability with which a link — and by extension, a segment or point — can be identified in the target database using one or two coordinates.

Using one coordinate, the search for the target link is straightforward: search for the point or segment in the target that is nearest to the source coordinate. Using two coordinates, which may in general be any two points along a source link, there are two ways to proceed:

1. Independent: Find the target point/segment nearest to the first source point, then by way of confirmation, the target point/segment nearest to the second source point. If the two source points snap to different target links, the transfer fails. [XSP2 does not constrain the two points to be on the same link; for test purposes our implementation *did* apply this restriction while generating sample points].
2. Simultaneous: Identify the target link with the lowest sum of distances between the two source points and the target points or segments nearest to them.

Pathologies can be constructed to show how each method can identify the wrong link. Each method can be evaluated based on corroboration by street name evidence where available. Although this test uses both coordinates and names, it is not a functional test of the {coordinates+names} implementation of XSP2; the purpose of name matching is only to validate the coordinate-based transfer.

Metrics

In general, street names may match, mismatch, or one or the other may be blank (due to the well understood difficulties inherent in name matching, this test may under-report results). There are 27 possible outcomes:

all streets match, one matches and two don't, … all streets mismatch. Cross street names must be cross-compared, i.e. from- must be compared against to- and vice versa.
Outcomes are scored as follows:

- HIT: Score 2 if the source and target street names match
- MISS: Score 0 if there is a mismatch between two non-blank names, OR if the source is blank but the target triad *can* be found elsewhere in the source file, OR the target is blank but the source triad *can* be found elsewhere in the target file.
- UNSURE: Score 1 otherwise.

On, From, To	Likely	Possible	Unlikely
0,0,0			•
0,1,0			•
0,2,0			•
0,0,1			•
0,1,1			•
0,2,1			•
0,0,2			•
0,1,2			•
0,2,2			•
1,0,0			•
1,1,0			•
1,2,0			•
1,0,1			•
1,1,1		•	
1,2,1		•	
1,0,2			•
1,1,2		•	
1,2,2		•	
2,0,0		•	
2,1,0		•	
2,2,0		•	
2,0,1		•	
2,1,1	•		
2,2,1	•		
2,0,2		•	
2,1,2	•		
2,2,2	•		

Table 10. Scoring of Transfers

Based on these scores, outcomes of the transfer are categorized using rules listed in Table 12. Corroboration is considered "likely" only if relatively strong scores are observed.

Sampling

Corroboration is examined for two samples:

(a) Machine-generated sample points, the same points used in Phase One
(b) Field-sampled points, the same points used in Phase One

Set (a) assigns equal weight to all links. While this is not relevant to ITS applications which tend to focus on major streets, it is appropriate to EMS applications. Point set (b) is biased in favor of ITS applications.

C. Transfer Using Names + Coordinates

This series of tests determines the effectiveness of XSP2 using street names in conjunction with coordinates. The process for identifying the target link uses (a) simple name matching, (b) fuzzy name matching, and (c) coordinate matching as tie-breaker and fall-back. Coordinates and names are meant to reinforce each other. Fallback to coordinates is undesirable in that there is no independent confirmation of transfer.

Metrics

The flowchart (Figure 12) shows how transfers are processed. Initially we seek a single perfect match between the source and destination names. Failing that, a limited set of fuzzy rules is applied. If multiple hits are encountered, the coordinate is introduced as a tie-breaker. If even the fuzzy match fails to find at least one point, the coordinate is used alone as a fallback. The metric for the test is the proportion of test points that falls into each of these outcome categories or "bins." Bins 3/4 and 7/8 are meant to trap cases distinguished by directionality, e.g. ambiguities such as "Birch between Ash and Ash." Bin 9 represents failure to resolve using names, resulting in fallback to coordinates. This is a critical bin since it measures the frequency with which names cannot be resolved at all.

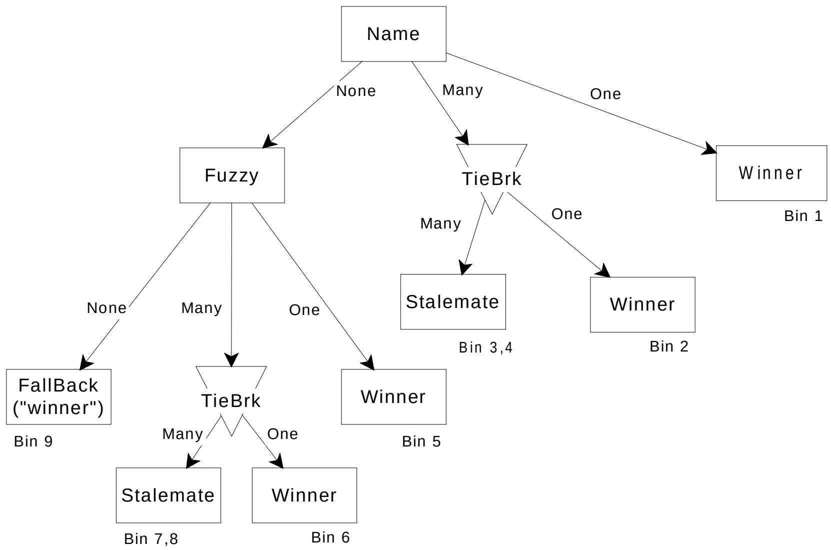

Figure 6. Name Matching and Coordinate Processing

Sampling

The sample points are the same as those used above. Processing is entirely automated.

D. Accuracy of Source Coordinate Transfer

This test measures the positional accuracy with which coordinate events are rendered in the target database. When a source point is snapped to the nearest target link (point or segment), the target link is tested to ensure that it is the intended link (see above), and positional error is recorded.

Metrics

The error value recorded is the Euclidean distance between the source and target point. Distances are broken down into three categories: 0–30 metres, 30–50 metres and more than 50 metres. A vector is constructed between the source and target points; vectors are mapped to show the distribution of error.

Sampling

The sample points are the same as those used above. Processing is entirely automated.

E. NQX Algorithm

As an afterthought to the original experimental plan, we implement a Nearest Qualifying Intersection (NQX) algorithm that searches intelligently for any occurrence of the required street names in the destination database, not necessarily as a triad of names associated with a single link. In effect, this approach forgives intervening streets in the destination. The method is particularly applicable to major-streets-only events typical of ITS.

The algorithm locates the two intersections {On, From} and {On, To} — blank names are not admitted at all. It examines all possible paths between these two intersections, constraining the search to links carrying the On-street name (in theory there should be only one path, but due to database errors and odd municipal practices, such as forking streets with the same name, multiple paths are often encountered). It selects the longest such path to be the destination street.

NQX is implemented in conjunction with other matching processes, in the sequence:

1. Exact match
2. NQX using exact match
3. Fuzzy match
4. NQX using fuzzy match

Metrics

Metrics are as above.

Sampling

Samples are picked as follows:

1. All streets in Santa Barbara county (as above)
2. Major streets only (cross streets are also restricted to major streets)
3. Combined urban area of Santa Barbara and Goleta

F. Accuracy of Offsets

Offset accuracy can be modeled as a function of link length and sinuosity. While this is relevant to the XSP/XSP2, it will be the focus of forthcoming lab and field tests of the Linear Referencing Profile.

Results

Results are tabulated in Appendix B. The principal observations are as follows:

A. Match to Self

This test is not performed in Phase Two.

B. Using Coordinates to Target a Link

Table 11 summarizes the results of Test Series B. On average, 11% of test points score “likely” matches, i.e. they snap to destination streets whose names compare favorably with the names in the source database. Most test points fall into the “possible” category, because (a) they match the wrong street in the right general area, or the wrong segment of the right street, due to coordinate error, or (b) confidence in name matching is diluted due to blanks and other name matching problems — this is unfortunately an inherent limitation of the test design. Results are far better for field-sampled points, which generally lie on named streets.

	Lab sample			Field sample		
	Mean	Min	Max	Mean	Min	Max
Likely	11%	2%	18%	24%	6%	46%
Possible	61%	52%	70%	26%	22%	31%
Unlikely	28%	18%	45%	50%	30%	65%

Table 11. Result Summary, Test Set B (single point transfer)

Appendix B also contains results from two-point independent and simultaneous transfers — used to transfer segments rather than points. The simultaneous method is invariably more successful, both for lab and field samples.

C. Transfer Using Names + Coordinates

Transfers are always "successful" using the method of names and coordinates. Two questions need to be asked to validate the transfers:

- In what percentage of cases does a transfer fall back to coordinates because a match is not possible using street names? This is addressed in Test Set C under "Bin 9," a reference to Bin 9 on the name processing flowchart (Figure 2).
- How far does the destination point lie from the source point? The name match could in some cases point to the wrong street, or if the destination database does not contain the road referenced in the source, transfer by coordinate may erroneously snap to the nearest available entity. These problems are addressed in Test Set D below.

	All Streets			Major Streets Only		
	Mean	Min	Max	Mean	Min	Max
Bin 9	35%	16%	45%	68%	38%	91%
Snap distance						
Median distance (m)	36.6	0.0	140.4	11.3	0.1	46.2
[0,30m)	66%	37%	96%	82%	52%	97%
[30,50m)	4%	0%	8%	5%	0%	15%
[50m and more)	30%	4%	58%	23%	3%	88%

Table 12 : Result Summary, Test Set C (lab generated single point transfer)

Detailed tabular results appear in Appendix B. Table 12 summarizes selected results pertaining to Test Sets C and D. The Bin 9 row indicates that 35% of all transfers (68% on major streets) fall back to coordinates. The high fallback for major streets is partly because aliases are not handled in this test set; they are examined in Test Set E, below.

Appendix B quotes results from two-point transfers using method C.

D. Accuracy of Source Coordinate Transfer

On studying the distance between source and destination points, we observe that distances below 30m are usually transfers made to the correct destination link, whereas distances above 50m are generally associated with errors.

For single point transfers, the median distance between source and destination points is about 35m. Two thirds of all points differ by less than 30m (probably hitting the correct link), one third by more than 50m (probably hitting the wrong link). The worst deviation observed in any test is 75 kilometres — clearly a referential error.

Since fallback points are snapped to the nearest arc without verifying street name, a high fallback rate in test set C is typically found in conjunction with an artificially lower median distance in D.

E. NQX Algorithm

The NQX approach produces the best results in the entire test series, with a mean fallback rate of 33% (compared with 35% not using the algorithm). For major streets, the fallback appears to rise from 68% to 72%; however, these numbers are *not* comparable because for the NQX algorithm, major street events are passed using only major cross streets; with other algorithms major street events may use minor cross streets. Limiting major streets to the Santa Barbara urban area, the average fallback is 51%, with a best score of 23%.

Although the magnitude of improvement due to NQX is disappointing, it is clear that this method is the most appropriate implementation of XSP, if only because transfers should not be confounded by the topology of intervening streets, the presence/absence of which are often matters of scale and interpretation. Based on manual checking of some results, it appears that when NQX fails, it is because of database errors, for example:

- Spelling disagreements in name records, between databases: Sargosso in database X, Sargoso in database Y
- Spelling errors or discontinuities in naming within a single database, due to which a path cannot be built connecting the two intersections points.

Some of these errors are easily fixed. We anticipate that intelligent name matching software could lower the fallback rate to the region of 10–20%.

Conclusions

Unfortunately, evaluation results of XSP cannot be summarized in a single figure or even a terse descriptive sentence. As we emphasized in the opening paragraphs, success is a function of numerous factors, from inherent profile effectiveness to municipal practices, database accuracy and quality of implementation. Since the XSP does not specify how a message is processed at the receiving end, implementation is inevitably a controlling factor in its success. Further, because there is no quality metadata explicitly associated with the message, there is potential for propagation of error. Because there are so many dimensions to the tests and results, each potential user group will need to focus on aspects appropriate to its needs.

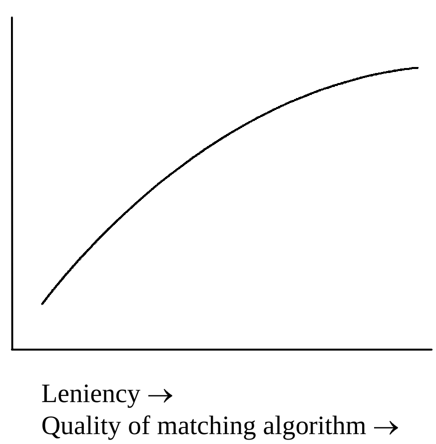

Figure 7: Success depends partly on implementation

For mission-critical applications such as EMS, the revised XSP (coordinates included) offers a measure of assurance, in that failure of transfer is clearly indicated by disagreement of coordinate and name components. Clearly EMS should take no comfort in the improvements in results due to “lenient” standards applied in Phase Two tests.

It must be emphasized that the bulk of the problem with low success rates is not the fault of the XSP specification, but is a reflection of database quality, particularly the high incidence of blanks, absence of alias fields in many databases, and non-standard name parsing and abbreviations. EMS agencies recognize the need for high quality data; in many areas they are the driving force behind municipal-level street database quality improvement initiatives. For best results EMS agencies must ensure that (a) they operate with reference to a single-source database as far as possible, and (b) they establish appropriate database quality control and testing measures.

The ITS industry expects far better success rates from a messaging standard than what has been achieved in these tests. The following are constructive recommendations for national-level activities that would lead to better results.

1. Obviously, the ideal long-term course of action is to re-survey the national street network to uniform quality standards. Efforts are already underway in piecemeal fashion. Several municipalities have integrated GIS programs in operation, with varying degrees of coordination between federal, state, county and private agencies. Outstanding hurdles are (a) the technical difficulty of finding a common quality standard that suits the needs of all stakeholders at a reasonable cost, and (b) management challenges to coordinate this activity at a national scale.

   Even if re-survey is publicly funded, commercial vendors will need to make substantial investments in data reorganization and conflation of nationwide databases. There are two shorter-term alternatives: standardization of databases, and the ITS Datum.

2. Standardization of databases: Messaging could be simplified if vendors would populate alias name fields, and comply with basic standards in street naming, in particular, highway and ramp nomenclature, field separation and abbreviation. Standardization of other aspects such as classification and inclusion, are desirable, but may not be readily achievable in the short term.

3. The ITS Datum (Siegel et al 1996) is a mid- to long-term strategy that could potentially

   - alleviate many current messaging problems,
   - provide an evolutionary framework for a high-quality national database, and
   - offer a mechanism for continuing update of highway-related coordinates and attributes, that would survive future construction and changes in geodetic datums.

Conceptual design of the ITS Datum is underway (Funk et al 1998; Church et al 1998); improvement in XSP success may be one of several measures of its ultimate effectiveness.

Funk C, K Curtin, M Goodchild, D Montello, V Noronha (1998) Formulation and test of a model of positional distortion fields. Third International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Quebec City [document available at www.ncgia.ucsb.edu/vital]

Church R, K Curtin, P Fohl, C Funk, M Goodchild, P Kyriakidis, V Noronha (1998) Positional distortion in geographic data sets as a barrier to interoperation. Proceedings, ACSM Baltimore [document available at www.ncgia.ucsb.edu/vital]

Goodwin CWH, D Siegel, S Gordon 1996. Location Reference Message Specification: Final Design. Task B: Spatial Data Interoperability Protocol For ITS Project, United States Department of Transportation, Federal Highway Administration Office of Safety and Traffic Operations, ITS Research Division, Contract 61-94-Y-00001, Draft, June 28, 1996

Haas RP, J Lau, C Goodwin, S Gordon (1998) Location Referencing Message Specification Report.

SAE (1998) Surface Vehicle Information Report — Location Referencing Message Specification. Society of Automotive Engineers, Information Report J2374

Siegel D, C Goodwin, S Gordon 1996. ITS Datum Final Design Report. Task C: Spatial Data Interoperability Protocol For ITS Project, United States Department of Transportation, Federal Highway Administration Office of Safety and Traffic Operations, ITS Research Division, Contract 61-94-Y-00001, Review Draft, June 28, 1996

VITAL 1997a. Interoperability of Map Databases — Development of Experimental Infrastructure. California Department of Transportation, Test Center for Interoperability, Interagency Agreement 65V250. Draft Final Report

VITAL 1997b. The Cross Streets LRMS Profile — Technical Evaluation. United States Department of Transportation, FHWA Contract DTFH61-91-Y-30066, Draft Phase I Report.

Appendix A — Phase One Outputs

Phase One result spreadsheets appear on the following pages

A.1 Transfer Results (Lab Tests)

A.2 Transfer Results (Field Tests)

Table A.1 Transfer Results (Lab Tests)

Source Database	**A**			**B**			**E**			**F**			**Mean**	**Min**	**Max**
Number of test points	10490	10490	10490	9053	9053	9053	6562	6562	6562	9251	9251	9251			
Number of valid triads	1153	1153	1153	1585	1585	1585	1875	1875	1875	1392	1392	1392			
Valid triad formation rate	11%	11%	11%	18%	18%	18%	29%	29%	29%	15%	15%	15%	18%	11%	29%
Target Database	**B**	**E**	**F**	**A**	**E**	**F**	**A**	**B**	**F**	**A**	**B**	**E**			
Ambiguities	15	20	8	11	18	15	18	18	12	10	21	11	15	8	21
Unambiguous hits															
Number of hits	261	361	135	280	346	537	360	315	98	125	523	92	286	92	537
Positional error, absolute offset															
Mean	20.0	69.9	18.4	28.8	73.4	6.6	72.1	83.7	80.9	20.1	5.4	93.9	47.8	5.4	93.9
Minimum	0.1	2.1	0.0	0.1	1.9	0.0	1.8	2.5	2.7	0.0	0.0	4.1	1.3	0.0	4.1
Maximum	600.0	788.4	226.2	1042.3	503.4	264.5	590.8	861.5	298.8	420.9	297.9	640.3	544.6	226.2	1042.3
Positional error, relative offset															
Mean	14.1	61.2	13.3	20.7	72.1	3.8	67.4	71.6	74.0	14.1	3.7	83.7	41.6	3.7	83.7
Minimum	0.1	1.7	0.0	0.1	1.9	0.0	0.5	1.7	2.9	0.1	0.0	3.8	1.1	0.0	3.8
Maximum	75.8	409.2	93.2	539.7	919.6	231.7	587.7	1046.9	283.5	53.7	167.6	651.1	421.6	53.7	1046.9
Hit rates															
Ambiguities	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Ambiguities, from valid triads	1%	2%	1%	1%	1%	1%	1%	1%	1%	1%	2%	1%	1%	1%	2%
Unambiguous hits	2%	3%	1%	3%	4%	6%	5%	5%	1%	1%	6%	1%	3%	1%	6%
Unambiguous, from valid triads	23%	31%	12%	18%	22%	34%	19%	17%	5%	9%	38%	7%	19%	5%	38%

Decimal values are metres

Table A.2 Transfer Results (Field Tests)

Source Database		**A**			**B**			**E**			**F**		**Mean**	**Min**	**Max**
Number of test points		54			54			54			54				
Number of valid triads		21			35			36			28		30	21	36
Valid triad formation rate		39%			63%			67%			52%		55%	39%	67%
Target Database	**B**	**E**	**F**	**A**	**E**	**F**	**A**	**B**	**F**	**A**	**B**	**E**			
Ambiguities	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1
Unambiguous hits															
Number of hits	4	9	3	4	6	19	9	7	3	3	19	3	7	3	19
Positional error, absolute offset															
mean	17.3	31.5	20.9	16.7	78.1	4.6	36.0	56.7	53.8	20.1	4.7	44.6	32.1	4.6	78.1
minimum	6.4	9.8	11.6	6.4	36.9	0.2	9.4	18.4	18.8	11.8	0.2	38.7	14.1	0.2	38.7
maximum	37.1	44.5	38.6	34.4	198.7	13.0	76.5	89.0	87.9	35.9	13.5	52.2	60.1	13.0	198.7
Positional error, relative offset															
mean	12.7	30.7	14.7	12.5	69.2	4.6	35.2	62.2	45.3	14.3	4.7	29.0	27.9	4.6	69.2
minimum	4.9	7.8	11.3	4.9	11.8	0.2	7.7	12.2	11.8	11.3	0.2	11.5	8.0	0.2	12.2
maximum	20.1	46.1	19.7	19.4	194.0	10.5	76.1	99.7	84.8	19.0	10.5	38.7	53.2	10.5	194.0
Hit rates															
Ambiguities	0%	0%	0%	2%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%
Ambiguities, from valid triads	0%	0%	0%	3%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
Unambiguous hits	7%	17%	6%	7%	11%	35%	17%	13%	6%	6%	35%	6%	14%	6%	35%
Unambiguous, from valid triads	19%	43%	14%	11%	17%	54%	25%	19%	8%	11%	68%	11%	25%	8%	68%

Decimal values are metres

Phase Two result spreadsheets appear on the following pages

Test Set B

Lab-generated sample points (about 1000)
B.1 Single point transfer
B.2 Two point independent transfer
B.3 Two point simultaneous transfer
Field sampled points (54)
B.4 Single point transfer
B.5 Two point independent transfer
B.6 Two point simultaneous transfer

Test Set C/D

All Streets

Lab-generated sample points (about 1000)
C.1 Single-point transfer
C.2 Two-point transfer
Field sampled points (54)
C.3 Single-point transfer
C.4 Two-point transfer

Major Streets

Lab-generated sample points (about 1000)
C.5 Single-point transfer
C.6 Two-point transfer

Test Set E (NQX Algorithm)

Single, lab-generated point transfer

E.1 All Streets
E.2 Major Streets
E.3 Santa Barbara/Goleta major urban streets

Table B.1 Lab generated, Single Point Transfer

Source Database	A			B			E			F			Mean	Min	Max
Number of test points	1049			905			656			925					
Target Database	B	E	F	A	E	F	A	B	F	A	B	E			
Match Statistics															
likely	15%	12%	6%	18%	10%	12%	18%	14%	3%	6%	11%	2%	11%	2%	18%
possible	66%	70%	66%	60%	61%	56%	56%	52%	52%	65%	65%	68%	61%	52%	70%
unlikely	19%	18%	28%	22%	29%	32%	26%	33%	45%	30%	24%	30%	28%	18%	45%
failed snap	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table B.2 Lab generated, Two Point Independent Transfer

Source Database	**A**			**B**			**E**			**F**			**Mean**	**Min**	**Max**
Number of test points	1049			905			656			925					
Target Database	**B**	**E**	**F**	**A**	**E**	**F**	**A**	**B**	**F**	**A**	**B**	**E**			
Match Statistics															
likely	14%	9%	6%	15%	9%	12%	12%	10%	2%	5%	11%	2%	9%	2%	15%
possible	60%	56%	61%	54%	50%	54%	33%	32%	31%	60%	63%	54%	51%	31%	63%
unlikely	15%	10%	23%	18%	18%	30%	14%	21%	26%	25%	23%	22%	20%	10%	30%
failed snap	12%	25%	10%	12%	23%	4%	41%	38%	40%	10%	3%	22%	20%	3%	41%

Table B.3 Lab generated, Two Point Simultaneous Transfer

Source Database	**A**			**B**			**E**			**F**			**Mean**	**Min**	**Max**
Number of test points	1049			905			656			925					
Target Database	**B**	**E**	**F**	**A**	**E**	**F**	**A**	**B**	**F**	**A**	**B**	**E**			
Match Statistics															
likely	16%	12%	6%	18%	12%	12%	19%	15%	4%	6%	11%	2%	11%	2%	19%
possible	66%	70%	67%	61%	61%	56%	56%	53%	52%	65%	65%	69%	62%	52%	70%
unlikely	18%	18%	27%	21%	27%	32%	25%	32%	44%	29%	24%	29%	27%	18%	44%
failed snap	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table B.4 Field sample, Single Point Transfer

Source Database	**A**			**B**			**E**			**F**			**Mean**	**Min**	**Max**
Number of test points	54			54			54			54					
Target Database	**B**	**E**	**F**	**A**	**E**	**F**	**A**	**B**	**F**	**A**	**B**	**E**			
Match Statistics															
likely	24%	35%	13%	24%	22%	46%	31%	20%	7%	13%	46%	6%	24%	6%	46%
possible	24%	24%	22%	28%	28%	22%	26%	26%	28%	24%	24%	31%	26%	22%	31%
unlikely	52%	41%	65%	48%	50%	31%	43%	54%	65%	63%	30%	63%	50%	30%	65%
failed snap	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table B.5 Field sample, Two Point Independent Transfer

Source Database		**A**			**B**			**E**			**F**		**Mean**	**Min**	**Max**
Number of test points		54			54			54			54				
Target Database	**B**	**E**	**F**	**A**	**E**	**F**	**A**	**B**	**F**	**A**	**B**	**E**			
Match Statistics															
likely	19%	22%	11%	17%	20%	46%	26%	19%	6%	11%	44%	4%	20%	4%	46%
possible	20%	15%	19%	19%	19%	20%	17%	19%	19%	17%	20%	19%	18%	15%	20%
unlikely	39%	30%	52%	35%	31%	31%	24%	35%	48%	41%	28%	39%	36%	24%	52%
failed snap	22%	33%	19%	30%	30%	2%	33%	28%	28%	31%	7%	39%	25%	2%	39%

Table B.6. Field sample, Two Point Simultaneous Transfer

Source Database	**A**			**B**			**E**			**F**			**Mean**	**Min**	**Max**
Number of test points	54			54			54			54					
Target Database	**B**	**E**	**F**	**A**	**E**	**F**	**A**	**B**	**F**	**A**	**B**	**E**			
Match Statistics															
likely	24%	31%	11%	22%	22%	46%	35%	19%	46%	13%	44%	4%	27%	4%	46%
possible	22%	28%	22%	30%	28%	22%	26%	26%	22%	26%	24%	31%	26%	22%	31%
unlikely	54%	41%	67%	48%	50%	31%	39%	56%	31%	61%	31%	65%	48%	31%	67%
failed snap	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table C.1. Lab generated, Single Point Transfer

Source Database	**A**			**B**			**E**			**F**			**Mean**	**Min**	**Max**
Number of test points	1049			905			656			925					
Target Database	**B**	**E**	**F**	**A**	**E**	**F**	**A**	**B**	**F**	**A**	**B**	**E**			
Distances															
0 - 30	71%	37%	80%	92%	41%	96%	55%	49%	51%	93%	86%	38%	66%	37%	96%
30 - 50	3%	6%	3%	1%	5%	0%	8%	8%	7%	2%	0%	5%	4%	0%	8%
> 50	26%	58%	17%	6%	54%	4%	37%	43%	42%	4%	14%	57%	30%	4%	58%
mean	860.6	1377.0	474.8	113.6	1148.9	217.0	1087.7	1431.5	1389.8	123.9	708.2	1589.9	876.9	113.6	1589.9
minimum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
maximum	37038.4	73456.6	48404.0	22257.5	30590.6	56032.0	61019.1	50822.9	44909.4	51007.9	37288.9	75133.8	48996.8	22257.5	75133.8
median	1.9	140.4	0.4	0.3	78.2	0.0	22.6	31.8	29.4	0.2	0.0	133.8	36.6	0.0	140.4
Match Resolution Category															
1	9%	8%	3%	9%	12%	19%	13%	16%	13%	3%	19%	10%	11%	3%	19%
2	31%	42%	21%	19%	35%	20%	26%	28%	28%	15%	28%	43%	28%	15%	43%
3	18%	7%	20%	21%	6%	22%	10%	7%	9%	22%	26%	9%	15%	6%	26%
4	8%	3%	9%	12%	1%	12%	5%	2%	3%	11%	11%	1%	6%	1%	12%
5	5%	5%	9%	7%	1%	2%	7%	1%	1%	11%	1%	1%	4%	1%	11%
6	0%	1%	0%	0%	0%	1%	0%	0%	1%	0%	1%	0%	0%	0%	1%
7	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	1%
8	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
9	28%	35%	38%	32%	45%	25%	40%	45%	45%	37%	16%	36%	35%	16%	45%

Table C.2. Lab generated, Two Point Transfer

Source Database	**A**			**B**			**E**			**F**			**Mean**	**Min**	**Max**
Number of test points	1049			905			656			925					
Target Database	**B**	**E**	**F**	**A**	**E**	**F**	**A**	**B**	**F**	**A**	**B**	**E**			
Distances															
0 - 30	68%	30%	78%	86%	35%	94%	36%	35%	35%	90%	86%	34%	59%	30%	94%
30 - 50	3%	7%	3%	3%	7%	1%	12%	9%	10%	3%	0%	6%	5%	0%	12%
> 50	29%	63%	19%	11%	59%	6%	52%	56%	55%	7%	14%	61%	36%	6%	63%
mean	865.1	1392.5	479.3	125.1	1181.6	220.8	1163.5	1483.7	1455.7	131.7	708.3	1616.9	902.0	125.1	1616.9
minimum	0.0	0.4	0.0	0.0	0.1	0.0	0.2	0.2	0.1	0.0	0.0	0.2	0.1	0.0	0.4
maximum	37366.6	73583.4	48538.2	22080.5	30600.7	55972.0	61552.0	51651.8	45096.8	50941.7	37070.0	75183.2	49136.4	22080.5	75183.2
median	4.0	344.0	0.8	0.4	112.4	0.0	44.3	64.1	53.4	0.3	0.0	175.6	66.6	0.0	344.0
Match Resolution Category															
1	9%	8%	3%	9%	12%	19%	13%	16%	13%	3%	19%	10%	11%	3%	19%
2	33%	43%	22%	21%	36%	22%	31%	31%	32%	16%	29%	43%	30%	16%	43%
3	22%	8%	24%	27%	6%	29%	9%	5%	7%	28%	32%	9%	17%	5%	32%
4	2%	1%	3%	3%	0%	3%	1%	0%	0%	3%	3%	0%	2%	0%	3%
5	5%	5%	9%	7%	1%	2%	7%	1%	1%	11%	1%	1%	4%	1%	11%
6	0%	1%	0%	0%	0%	1%	0%	0%	1%	0%	1%	1%	0%	0%	1%
7	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	1%
8	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
9	28%	35%	38%	32%	45%	25%	40%	45%	45%	37%	16%	36%	35%	16%	45%

Table C.3. Field sample, Single Point Transfer

Source Database	A			B			E			F			**Mean**	**Min**	**Max**
Number of test points	54			54			54			54					
Target Database	**B**	**E**	**F**	**A**	**E**	**F**	**A**	**B**	**F**	**A**	**B**	**E**			
Distances															
0 - 30	81%	72%	85%	89%	76%	100%	74%	72%	70%	89%	98%	76%	82%	70%	100%
30 - 50	6%	19%	7%	6%	13%	0%	19%	15%	13%	7%	0%	13%	10%	0%	19%
> 50	13%	9%	7%	6%	11%	0%	7%	13%	17%	4%	2%	11%	8%	0%	17%
mean	439.7	27.2	15.3	12.6	21.8	2.2	33.1	436.4	1745.7	8.3	630.7	22.9	283.0	2.2	1745.7
minimum	0.2	0.7	0.0	0.2	0.2	0.0	1.0	0.2	1.3	0.0	0.0	0.2	0.3	0.0	1.3
maximum	22229.4	155.1	90.7	59.0	91.0	16.7	574.6	22228.3	36765.7	91.0	33953.3	91.0	9695.5	16.7	36765.7
median	8.8	22.3	7.9	7.3	17.5	1.1	20.2	19.0	22.0	0.1	1.0	20.3	12.3	0.1	22.3
Match Resolution Category															
1	17%	24%	9%	19%	22%	59%	24%	28%	26%	11%	63%	24%	27%	9%	63%
2	4%	4%	0%	0%	0%	0%	4%	2%	6%	0%	2%	2%	2%	0%	6%
3	0%	0%	0%	2%	2%	2%	7%	0%	0%	0%	2%	2%	1%	0%	7%
4	0%	6%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	6%
5	31%	22%	43%	30%	0%	2%	22%	0%	0%	41%	2%	2%	16%	0%	43%
6	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
7	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%	2%
8	0%	0%	0%	2%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%	2%
9	48%	44%	48%	48%	76%	37%	43%	70%	69%	44%	31%	70%	52%	31%	76%

Table C.4. Field sample, Two Point Transfer

Source Database	A			B			E			F			Mean	Min	Max
Number of test points	54			54			54			54					
Target Database	**B**	**E**	**F**	**A**	**E**	**F**	**A**	**B**	**F**	**A**	**B**	**E**			
Distances															
0 - 30	72%	56%	76%	76%	57%	98%	57%	59%	57%	70%	96%	54%	69%	54%	98%
30 - 50	7%	20%	9%	13%	17%	2%	26%	17%	15%	19%	0%	17%	13%	0%	26%
> 50	20%	24%	15%	11%	26%	0%	17%	24%	28%	11%	4%	30%	17%	0%	30%
mean	444.1	34.2	21.5	19.9	30.5	2.5	36.7	441.3	1757.4	21.2	632.8	36.2	289.9	2.5	1757.4
minimum	0.7	2.3	0.0	0.6	0.3	0.0	1.9	1.8	1.5	0.4	0.0	1.1	0.9	0.0	2.3
maximum	22130.8	203.5	199.8	193.4	156.3	28.7	586.2	22113.8	37250.4	193.5	33982.2	242.5	9773.4	28.7	37250.4
median	10.0	25.1	8.9	11.6	23.8	1.3	22.8	20.3	23.7	10.5	1.4	25.6	15.4	1.3	25.6
Match Resolution Category															
1	17%	24%	9%	19%	22%	59%	24%	28%	26%	11%	63%	24%	27%	9%	63%
2	4%	4%	0%	0%	2%	0%	4%	2%	6%	0%	2%	4%	2%	0%	6%
3	0%	4%	0%	2%	0%	2%	7%	0%	0%	0%	2%	0%	1%	0%	7%
4	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%
5	31%	22%	43%	30%	0%	2%	22%	0%	0%	41%	2%	2%	16%	0%	43%
6	0%	0%	0%	2%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%	2%
7	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%	2%
8	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
9	48%	44%	48%	48%	76%	37%	43%	70%	69%	44%	31%	70%	52%	31%	76%

Table C.5. Lab generated, Single Point Transfer, Major Streets Only

Source Database	**A**				**B**				**D**				**E**				**F**				**Mean**	**Min**	**Max**
Number of test points	180				118				105				162				53						
Target Database	**B**	**D**	**E**	**F**	**A**	**D**	**E**	**F**	**A**	**B**	**E**	**F**	**A**	**B**	**D**	**F**	**A**	**B**	**D**	**E**			
Distances																							
0 - 30	82%	52%	71%	87%	92%	73%	69%	97%	86%	83%	75%	84%	69%	70%	45%	64%	94%	75%	83%	72%	76%	45%	97%
30 - 50	2%	1%	5%	1%	3%	3%	8%	0%	5%	5%	8%	6%	7%	6%	8%	3%	2%	2%	2%	8%	4%	0%	8%
> 50	17%	48%	24%	12%	4%	24%	22%	3%	10%	12%	17%	10%	24%	25%	47%	33%	4%	23%	15%	21%	20%	3%	48%
mean	1228.3	4543.5	529.4	1551.4	15.5	1660.1	270.3	944.5	1411.0	1834.3	59.4	1431.6	207.0	416.9	2517.4	1927.0	343.9	2639.7	1699.6	376.5	1280.4	15.5	4543.5
minimum	0.0	0.1	0.1	0.0	0.0	0.2	0.1	0.0	0.1	0.3	0.1	0.3	0.2	0.2	0.3	0.2	0.0	0.0	0.0	0.1	0.1	0.0	0.3
maximum	35362	55965	13240	55564	524	56072	12161	55671	43247	53318	1793	53318	8590	20727	74615	45241	17751	40980	28713	5061	33896	524	74615
median	2.9	24.5	11.2	1.4	2.6	11.7	16.5	0.3	9.1	12.9	12.0	11.9	11.6	13.1	46.2	15.2	1.0	0.1	11.4	11.1	11.3	0.1	46.2
Match Resolution Category																							
1	7%	3%	11%	0%	8%	15%	16%	33%	7%	15%	7%	13%	11%	9%	6%	7%	0%	15%	0%	4%	9%	0%	33%
2	17%	18%	15%	10%	2%	6%	4%	6%	6%	5%	2%	4%	12%	9%	10%	17%	4%	25%	9%	9%	9%	2%	25%
3	18%	0%	18%	11%	8%	0%	4%	12%	1%	0%	2%	2%	15%	8%	1%	10%	11%	17%	0%	2%	7%	0%	18%
4	7%	1%	6%	4%	5%	2%	0%	6%	0%	0%	0%	0%	4%	0%	0%	1%	4%	0%	0%	2%	2%	0%	7%
5	13%	8%	7%	18%	12%	1%	0%	1%	7%	0%	0%	1%	6%	1%	0%	3%	2%	0%	0%	0%	4%	0%	18%
6	0%	0%	2%	0%	0%	0%	0%	1%	0%	0%	0%	0%	1%	0%	1%	4%	0%	0%	0%	0%	0%	0%	4%
7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	1%
8	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%
9	38%	70%	42%	57%	64%	76%	75%	42%	80%	80%	90%	80%	52%	73%	83%	59%	79%	43%	91%	83%	68%	38%	91%

Table C.6. Lab generated, Two Point Transfer, Major Streets Only

Source Database	A				B				D				E				F				Mean	Min	Max
Number of test points	180				118				105				162				53						
Target Database	**B**	**D**	**E**	**F**	**A**	**D**	**E**	**F**	**A**	**B**	**E**	**F**	**A**	**B**	**D**	**F**	**A**	**B**	**D**	**E**			
Distances																							
0 - 30	78%	49%	65%	85%	76%	72%	59%	93%	43%	38%	36%	34%	38%	47%	38%	36%	85%	72%	75%	64%	59%	34%	93%
30 - 50	3%	1%	7%	2%	9%	3%	10%	1%	6%	11%	10%	11%	9%	7%	9%	5%	6%	2%	2%	8%	6%	1%	11%
> 50	18%	49%	28%	13%	14%	25%	31%	6%	51%	50%	54%	54%	54%	46%	54%	59%	9%	26%	23%	28%	35%	6%	59%
mean	1232.7	4541.3	532.4	1559.2	31.6	1669.7	280.5	950.1	2033.8	2353.2	718.9	2124.8	346.8	502.1	2562.6	2059.4	355.3	2671.1	1724.8	391.8	1432.1	31.6	4541.3
minimum	0.0	0.2	0.5	0.0	0.0	0.5	1.0	0.0	0.8	0.9	1.4	1.5	0.5	0.3	1.5	0.3	0.0	0.0	0.2	2.1	0.6	0.0	2.1
maximum	35422	55658	13268	55257	520	56109	12202	55708	38635	48923	7618	48923	8415	20757	74789	44927	18137	41264	28689	5118	33517	520	74789
median	3.7	24.0	13.3	1.7	5.0	12.8	19.9	0.6	56.1	54.3	57.0	63.9	41.9	27.1	51.5	61.2	3.9	0.3	22.7	31.0	27.6	0.3	63.9
Match Resolution Category																							
1	7%	3%	11%	0%	8%	15%	16%	33%	7%	15%	7%	13%	11%	9%	6%	7%	0%	15%	0%	4%	9%	0%	33%
2	19%	18%	17%	12%	3%	6%	4%	7%	6%	5%	4%	5%	20%	14%	10%	24%	4%	25%	9%	9%	11%	3%	25%
3	22%	0%	19%	13%	10%	0%	4%	16%	1%	0%	0%	1%	9%	4%	1%	3%	15%	17%	0%	2%	7%	0%	22%
4	1%	1%	2%	0%	2%	2%	0%	1%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	2%	1%	0%	2%
5	13%	8%	7%	18%	12%	1%	0%	1%	7%	0%	0%	1%	6%	1%	0%	3%	2%	0%	0%	0%	4%	0%	18%
6	0%	0%	2%	0%	0%	0%	0%	1%	0%	0%	0%	0%	1%	0%	1%	4%	0%	0%	0%	0%	0%	0%	4%
7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
8	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%
9	38%	70%	42%	57%	64%	76%	75%	42%	80%	80%	90%	80%	52%	73%	83%	59%	79%	43%	91%	83%	68%	38%	91%

Table E.1. Santa Barbara County, All Streets

Source Database	**A**			**B**			**E**			**F**			**Mean**	**Min**	**Max**
Number of test points	1049			905			656			925					
Target Database	**B**	**E**	**F**	**A**	**E**	**F**	**A**	**B**	**F**	**A**	**B**	**E**			
Distances															
0 - 30	69%	36%	80%	92%	41%	96%	54%	48%	50%	93%	86%	38%	65%	36%	96%
30 - 50	3%	6%	3%	1%	6%	0%	8%	8%	7%	2%	0%	5%	4%	0%	8%
> 50	28%	58%	17%	7%	54%	4%	38%	44%	42%	5%	14%	57%	31%	4%	58%
mean	1130.3	1411.9	471.9	106.7	1149.0	217.0	1171.9	1432.3	1390.9	128.4	708.5	1590.4	909.1	106.7	1590.4
minimum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
maximum	81499.8	73456.6	45270.9	22257.5	30590.6	56032.0	61019.1	50822.9	44909.4	51007.9	37288.9	75133.8	52440.8	22257.5	81499.8
median	2.1	147.0	0.4	0.3	78.7	0.0	23.9	32.2	29.9	0.2	0.0	134.1	37.4	0.0	147.0
Match Resolution Category															
1	8%	8%	3%	8%	12%	19%	13%	16%	13%	3%	19%	10%	11%	3%	19%
2	34%	43%	22%	19%	36%	20%	29%	30%	30%	16%	28%	43%	29%	16%	43%
3	19%	7%	21%	21%	6%	22%	11%	7%	9%	22%	26%	9%	15%	6%	26%
4	8%	3%	9%	12%	1%	12%	6%	2%	4%	12%	11%	2%	7%	1%	12%
5	5%	5%	9%	7%	1%	2%	7%	1%	1%	11%	1%	1%	4%	1%	11%
6	0%	1%	1%	1%	0%	1%	0%	0%	1%	1%	1%	1%	1%	0%	1%
7	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	1%
8	0%	0%	0%	0%	0%	0%	1%	0%	0%	1%	0%	0%	0%	0%	1%
9	25%	33%	36%	32%	44%	25%	34%	43%	42%	34%	15%	35%	33%	15%	44%

Table E.2. Santa Barbara County, Major Streets Only

Source Database	A				B				D				E				F				Mean	Min	Max
Number of test points	154				118				99				137				53						
Target Database	**B**	**D**	**E**	**F**	**A**	**D**	**E**	**F**	**A**	**B**	**E**	**F**	**A**	**B**	**D**	**F**	**A**	**B**	**D**	**E**			
Distances																							
0 - 30	85%	66%	75%	92%	90%	79%	78%	100%	91%	91%	77%	90%	73%	75%	47%	73%	91%	87%	91%	77%	81%	47%	100%
30 - 50	5%	1%	6%	2%	3%	3%	8%	0%	2%	4%	8%	5%	13%	12%	4%	12%	2%	2%	2%	8%	5%	0%	13%
> 50	10%	32%	19%	6%	8%	19%	14%	0%	7%	5%	15%	5%	14%	13%	49%	15%	8%	11%	8%	15%	14%	0%	49%
mean	139.8	1035.2	226.7	76.9	383.1	216.4	28.0	1.3	1488.8	1415.9	30.6	1415.7	61.5	29.1	1179.3	123.5	532.6	845.8	151.1	34.3	470.8	1.3	1488.8
minimum	0.0	0.2	0.1	0.0	0.0	0.2	0.1	0.0	0.2	0.4	0.1	0.4	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.4
maximum	4620	18234	8878	5215	16704	3878	527	24	45420	55537	689	55537	3804	407	15541	5608	17751	40980	6137	766	15313	24	55537
median	2.6	11.9	11.5	1.0	3.1	10.5	14.0	0.2	8.4	9.6	11.9	9.1	15.2	15.3	47.2	14.2	1.6	0.0	10.8	11.0	10.5	0.0	47.2
Match Resolution Category																							
1	4%	5%	5%	0%	0%	0%	2%	3%	2%	1%	1%	2%	1%	1%	1%	1%	0%	6%	0%	0%	2%	0%	6%
2	23%	5%	19%	3%	15%	7%	12%	19%	13%	15%	2%	8%	13%	16%	5%	12%	8%	26%	2%	0%	11%	0%	26%
3	6%	0%	6%	2%	2%	0%	0%	3%	0%	0%	0%	0%	4%	2%	0%	2%	6%	4%	0%	0%	2%	0%	6%
4	6%	4%	6%	1%	8%	8%	1%	14%	9%	8%	1%	9%	12%	11%	5%	7%	0%	0%	0%	2%	6%	0%	14%
5	3%	2%	1%	5%	1%	2%	1%	0%	1%	1%	0%	1%	2%	0%	0%	0%	0%	0%	0%	0%	1%	0%	5%
6	5%	3%	5%	24%	9%	0%	0%	3%	3%	0%	0%	0%	7%	0%	0%	1%	15%	0%	0%	2%	4%	0%	24%
7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
8	8%	5%	6%	14%	8%	0%	0%	1%	3%	0%	0%	0%	9%	0%	0%	1%	8%	0%	0%	2%	3%	0%	14%
9	45%	77%	50%	51%	57%	84%	85%	58%	69%	75%	96%	80%	53%	69%	89%	76%	64%	64%	98%	94%	72%	45%	98%

Table E.3. City of Santa Barbara, Major Streets Only

Source Database	**A**				**B**				**D**				**E**				**F**				**Mean**	**Min**	**Max**
Number of test points		144				117				98				136				28					
Target Database	**B**	**D**	**E**	**F**	**A**	**D**	**E**	**F**	**A**	**B**	**E**	**F**	**A**	**B**	**D**	**F**	**A**	**B**	**D**	**E**			
Distances																							
0 - 30	86%	97%	73%	85%	94%	93%	79%	99%	97%	90%	81%	87%	66%	67%	77%	68%	82%	100%	93%	75%	84%	66%	100%
30 - 50	6%	1%	13%	6%	4%	6%	9%	0%	1%	9%	13%	11%	17%	15%	9%	14%	11%	0%	4%	11%	8%	0%	17%
> 50	8%	1%	14%	9%	2%	1%	11%	1%	2%	1%	6%	2%	17%	18%	15%	18%	7%	0%	4%	14%	8%	0%	18%
mean	18.4	10.6	32.7	18.3	10.4	11.1	23.6	3.1	12.9	15.2	20.2	15.8	53.7	45.1	38.8	45.0	71.2	2.1	12.1	29.0	24.5	2.1	71.2
minimum	0.0	0.0	0.0	0.3	0.0	0.1	0.1	0.0	0.4	0.1	0.0	0.6	0.3	0.3	0.1	0.6	0.1	0.0	0.5	0.2	0.2	0.0	0.6
maximum	197	231	705	197	77	75	153	60	494	61	280	62	1630	1583	1587	1583	838	15	32	226	504	15	1630
median	11.1	6.5	19.4	10.8	6.0	8.1	15.9	1.6	5.9	12.0	14.3	12.2	22.1	20.5	14.0	21.4	10.3	0.6	10.8	19.4	12.1	0.6	22.1
Match Resolution Category																							
1	6%	6%	3%	2%	9%	17%	3%	17%	3%	6%	7%	7%	1%	1%	1%	1%	11%	21%	4%	7%	7%	1%	21%
2	14%	15%	17%	1%	9%	13%	19%	18%	7%	13%	7%	12%	9%	10%	9%	9%	4%	11%	7%	0%	10%	0%	19%
3	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	2%	1%	1%	0%	0%	0%	11%	0%	0%	0%	1%	0%	11%
4	17%	14%	12%	3%	15%	37%	10%	40%	21%	34%	8%	36%	12%	9%	9%	8%	4%	14%	11%	0%	16%	0%	40%
5	5%	5%	7%	9%	8%	0%	0%	1%	2%	0%	0%	0%	6%	1%	1%	1%	0%	0%	0%	0%	2%	0%	9%
6	8%	8%	7%	22%	5%	0%	0%	0%	4%	0%	0%	0%	15%	0%	0%	1%	7%	0%	0%	0%	4%	0%	22%
7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%
8	19%	17%	28%	31%	24%	0%	0%	1%	14%	0%	0%	0%	32%	0%	0%	1%	14%	0%	0%	0%	9%	0%	32%
9	31%	35%	26%	32%	29%	33%	68%	23%	48%	47%	76%	44%	23%	79%	80%	78%	50%	54%	79%	93%	51%	23%	93%