Coordinate Systems and Terrain Reusability

Susanne Z. Riehemann
SRI International
333 Ravenswood Ave
Menlo Park, CA 94025
susanne.riehemann@sri.com

Keywords:
coordinate systems, position accuracy, plate tectonics, terrain reusability, live training

ABSTRACT: *This paper is about a coordinate system problem for the "live" aspect of LVC training that impacts terrain reusability. Commonly used commercial software for producing correlated terrains has an outdated definition of datums that assumes that NAD83 and WGS84 are still equivalent, and does not take into account the newer realizations of these, or the fact that there have been significant plate tectonic and other changes of the earth since 1984. This results in a total overall difference of up to 1.7 meters between a synthetic 3D terrain made from NAD83 survey data and satellite imagery, and the live tracking data from a training exercise, which is in the latest realization of WGS84 at the epoch of the exercise. There is no available GIS software that includes an implementation of the latest realization of WGS84 (G1150) or time based transformations that can be used for imagery or elevation data. So even when pre-processing data, there is a remaining terrain error that varies depending on the plate tectonic velocity of the area of interest - currently about 65cm total error in some parts of California. The paper discusses two possible solutions to this problem, and their implications for terrain reusability. In particular, if a terrain is transformed to match the live tracking data for a particular event, then the terrain is specific to that epoch and not reusable for later events without loss of accuracy. A better solution would be to agree on a standard epoch for terrains, and use the Horizontal Time Dependent Positioning (HTDP) software to transform the tracking data to match that epoch. It is also important to ensure that all WGS84 source data includes the epoch as part of the metadata.*

1. Introduction

The coordinate system used by GPS is moving underneath our feet. More accurately, due to plate tectonics, points on the earth's surface move with respect to this reference system. If this is not taken into account when producing synthetic 3D terrains, it results in problems for the "live" aspect of Live-Virtual-Constructive (LVC) training, and for any other application that uses data with mixed datums and requires accurate positioning.

Correlated terrain is an essential part of making LVC exercises work; for example avoiding fair fight issues. Early in the Joint Training Experimentation Program (JTEP), SRI appreciated the requirement for correlated terrain between LVC federates to support tightly coupled interactions, such as cross-attached entities and direct-fire engagements. As a result of this requirement, SRI began to develop highly accurate correlated synthetic terrain databases for all of the LVC federates participating in an exercise [1], [2]. Because most of the JTEP exercises involved live entities, the live terrain became the reference to which terrain databases developed for virtual and constructive federates were correlated.

The synthetic terrain is a model of the real world terrain. It is based on digital elevation data and satellite imagery, and includes ground relief, such as hills and valleys, and an image of features overlaid on the ground, similar to Google Earth. To this basic terrain, JTEP, and subsequently the Exportable Combat Training Capability (XCTC) program, adds photo-realistic geo-referenced buildings and other features relevant to the exercise, in order to create a truly 3D geo-referenced synthetic terrain database. Live entities in the training event are tracked by GPS, and corresponding avatars move through the synthetic terrain. This allows for After Action Reviews (AARs) that can show each significant event from any point of view desired, and that can include other information, such as name labels or shot lines.

Live GPS tracking for ground based entities like vehicles and people can be accurate to about 1 meter of uncertainty in horizontal positioning if differential GPS is used and ephemeris, ionospheric, tropospheric and earth tide effects are taken into account [3]. WAAS-enabled GPS receivers can get close to this accuracy in low multipath environments [4]. The important features in the training area of a live exercise, such as the corners of buildings, doors, windows, fences, etc., can be surveyed with an accuracy of about 2cm when correctly using state-of-the-art GPS equipment and Online Positioning User Service (OPUS) corrections [5].

Because the error in the locations of the surveyed terrain features is negligible, the total overall uncertainty about the position of an entity with respect to the simulated terrain should be the same as the tracking uncertainty, so currently about 1 meter.

While producing terrains for the XCTC and JTEP programs we discovered a systematic shortcoming in the datum implementations of current software. It can cause the overall error to be almost three times as large as it currently must be, given the tracking uncertainty. This paper describes the problem and suggests both a workaround and a longer-term solution for it.

A note on confusing terminology is necessary at this point. The term "WGS84" (World Geodetic System 1984) is often used ambiguously for at least five related things. It is sometimes used as the name of a particular reference ellipsoid (an approximation of the earth surface). Most commonly it is the name of a datum, which defines the position of the reference ellipsoid with respect to the center and the surface of the earth, and which is used for a variety of coordinate systems including projected ones. WGS84 is also often used as the name of the geodetic (also called geographic) coordinate system in which latitude is the angle (in degrees) formed by the normal to the ellipsoid and the equatorial plane, and longitude is the angle with respect to the reference meridian. There is also a WGS84 geocentric cartesian ECEF (Earth Centered Earth Fixed) coordinate system, which is what GPS uses, although the coordinates are usually converted to geodetic coordinates for the user. Finally, all of the above plus more (a Geoid, an Earth Gravitational Model, ...) make up the WGS84 system as a whole.

The problem at hand is common to all coordinate systems based on the WGS84 datum, whether they are geocentric, geodetic or projected.

2. The Problem

2.1 Outdated datums in terrain software

SRI uses industry-standard commercial software to produce geo-referenced 3D terrains for live entities, and to produce other correlated terrains for a variety of LVC systems, including JCATS, OneSAF, etc. This terrain generating software includes implementations of datums that are outdated, and assumes that NAD83 (North American Datum 1983) and WGS84 (World Geodetic System 1984) are still equivalent, as they were in 1984. But there are newer realizations of these two reference systems. They were updated since the advent of GPS to be more precise (e.g. about the exact center of the earth and reference stations on the surface of the earth), and they now use different ellipsoids. The most relevant newer realizations are NAD83 (CORS96) and WGS84 (G1150), although there are others to consider. Due to these new realizations, and, more significantly, due to plate tectonic and other earth changes, NAD83 and WGS84 are no longer the same.

NAD83 is a datum that is relative to reference stations on the North American plate at a particular epoch (i.e. at a particular point in time) [6]. It is used for construction work, surveying property lines, and other legal purposes, where it is important that the coordinates do not change over time. This datum also facilitates more traditional surveying approaches based on previously surveyed benchmarks. Because surveyors are used to working with this datum, survey data is usually provided in it. The aerial imagery and elevation data that the synthetic terrain is based on is also frequently provided in a projected State Plane coordinate system with a NAD83 datum.

GPS coordinates, including those from live entity tracking during an exercise event, are WGS84 coordinates. WGS84 is now tied to the International Terrestrial Reference Frame (ITRF) [7]. Because it has to be usable for the GPS satellites and for positions all over the earth, latitudes and longitudes are relative to a global coordinate system that is fixed with respect to the earth's center, the rotational axis, and the IERS Reference Meridian (IRM). Note that the IRM is not the same as the Greenwich "Prime Meridian', which is more than a hundred meters further west. In fact, the IRM does not correspond to any particular fixed location, but is a computed average based on many reference stations, to ensure that there is no net horizontal tectonic motion with respect to the system as a whole. The coordinates for these stations are updated to be more precise for each new realization, and include velocities. As a result, WGS84 coordinates are not fixed with respect to any particular points on the North American Plate or anywhere on the earth's surface. Due to plate tectonics, the WGS84 coordinates of a particular location change over time.

As an example, the exact WGS84 coordinate of the geodetic survey plaque on top of Mount Diablo in California changes over time, as the North American Plate, on which the survey plaque is located, drifts with respect to the WGS84 datum. If precise latitude and longitude values were engraved on the plaque years ago, those do not correspond to the WGS84 coordinates one would get from a precise GPS survey today. Similarly, there is no place where one could put a precise "equator" line monument that would stay accurate enough that visitors can stand with one foot in each hemisphere, although there are several places that make such claims.

Figures 1 and 2 illustrate the coordinate system staying fixed while the continents on the earth surface drift. Someone might place a plaque at the black point that is exactly at $40°\mathrm{N}, 120°\mathrm{W}$ in Figure 1, but then the tectonic plate on which the plaque is located will drift, and the coordinate will no longer be precisely accurate. For illustrative purposes, to show continents moving while the latitude and longitude lines stay fixed, Figure 2 shows so much exaggerated drift that the point is no longer in Northern California but instead in San Diego.

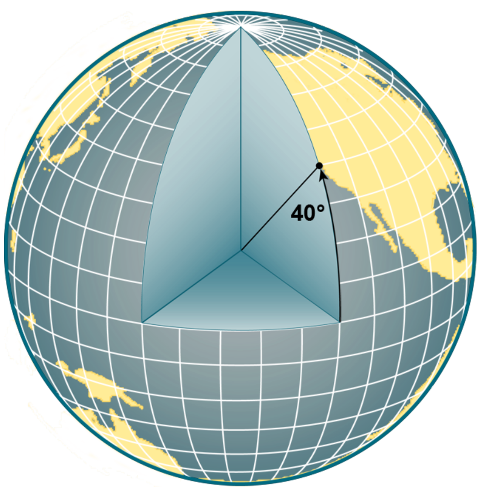

Figure 1

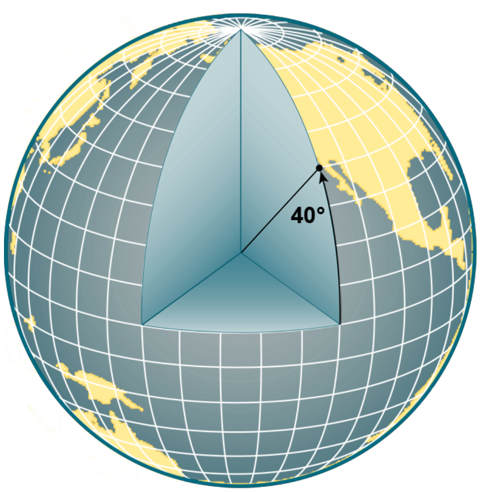

Figure 2

When geo-referenced simulated terrain is developed based on NAD83 data, these differences can be significant - as much 1.7 meters between the location of a feature in the simulated terrain and the same location in the real world at the time of the exercise, as determined by the live GPS tracking data. Most of that difference is due to plate tectonic movements since 1984. This portion gets worse every year. The remainder is due to the changes in the newer datum realizations like WGS84 (G1150). The exact amount of this difference varies in different parts of the US. The 1.7 meter figure is an example for a reference station at the Vandenburg Air Force Base on the Pacific Plate in California. In Florida the difference is about 65 cm.

As stated previously, the total difference between a tracked entity and the synthetic terrain should only be as large as the tracking uncertainty - about 1 meter. This means that the 1.7 meter inaccuracy of the terrain introduced by the software's outdated datum definitions can almost triple the overall positional error to 2.7 meters in the worst case scenario when the two errors happen to add to each other. The root sum square of that error is about 2 meters.

Depending on the training purpose, the difference between a 1 meter error and a 2.7 meter error can significantly impact the value of the AAR. Given that a person is about 0.5m wide, if one can see a person in the real world from some given location, there is a high probability that from the same location in the synthetic terrain, one can see at least part of the avatar for that person if the error is 1m. But with a 2.7 meter error, the avatar could be next to a door on one side of a building, when in reality the person is around a corner on a windowless side of the building. This kind of error can make the difference between having a clear line-of-sight (LOS) and not being able to see someone who is obscured by an incorrectly placed terrain feature (see Figure 3 and Figure 4). It can result in fair fight issues between live and virtual players, or between two virtual systems with terrains that differ in this regard.

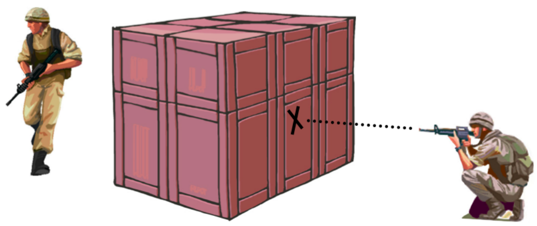

Figure 3

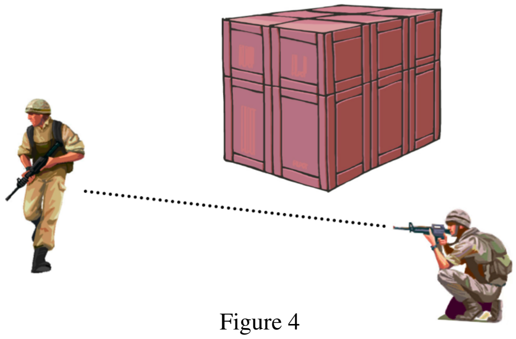

Figure 4

2.2 Limitations of GIS Pre-Processing

It is possible to use GIS software to pre-process the inputs to the terrain production software to eliminate part of this error. Specifically, software transformations can move the NAD83 survey data and imagery to the WGS84 (G873) realization, which has an epoch of 1996.0. However, this software does not yet include an implementation of the latest realization of WGS84 (G1150), nor does it include a time based translation like the HTDP (Horizontal Time Dependent Positioning) software developed by the National Geodetic Survey (NGS) in 1992 [8]. The HTDP software can be used only for individual coordinates and not for imagery or elevation data. For this reason, even when using GIS pre-processing tools, there is a remaining error. The exact amount of this error varies depending on the velocity, due to plate tectonics, of the area of interest. In the contiguous United States it is worst in California, where some locations on the Pacific Plate have velocities of 5cm/year, resulting in a total shift of 65 cm since 1996. In some parts of the world, the velocity is greater than 7cm/year [9]. Even in the middle of the North American Plate (e.g. in Michigan), it can still be over 2cm/year, resulting in a 26cm shift since 1996.

2.3 Incomplete Metadata

There is also a problem with older WGS84 data. The GIS software does not provide transformations for different epochs of WGS84. In addition, many providers of source data are not aware of this problem and do not specify a date. And when there is a date, no distinction is made between the date the imagery was taken and the epoch with respect to which the imagery was georeferenced. In theory, even new imagery could be referenced with respect to WGS data that is 25 years old, and could have the same 1.7 meter error as NAD83 data does. Given that current imagery is often 25cm resolution, sometimes even 10cm resolution, the missing metadata can significantly compromise the accuracy of these high resolution data sets.

Similarly, different realizations of NAD83 are not always distinguished in the metadata. The 1.7m error in Vandenberg, CA is based on the difference between WGS84 (G1150) (EPOCH 09-01-2009) and NAD83 (CORS96) (EPOCH 01-01-2002), which is what one would get from a surveyor who uses GPS and a standard OPUS correction. However, there are other realizations of NAD83. NAD83 (NSRS2007) (National Spatial Reference System) is used by some RTK (Real-Time Kinematic) networks for surveying. On the other hand, older realizations cause even larger errors. NAD83 (CORS96) positions were originally used with an epoch of 1997, which results in an error of almost 1.9m for Vandenberg. The NAD83 (HARN) (High Accuracy Reference Network) realizations differ on a state-by-state basis. If original NAD83 (1986) benchmarks are used, the error is even worse. Again the main problem is that these realizations of NAD83 are not always specified in the metadata, so it is impossible to know how to properly transform to WGS84. A secondary problem lies in the fact that there is no software to transform directly from all of these into the current epoch of WGS84.

Although this problem has been uncovered as a result of our development of highly accurate geo-referenced synthetic terrain, the issue potentially has implications for a wide variety of applications of GPS-based tracking, robot navigation, surveying, and other data collection and use. Our society is moving in a direction where many people produce GIS data. GPS is integrated into many modern devices such as cell phones, cameras, and cars. Many consumer GPS devices are already WAAS enabled. Potentially corrections could also be broadcast over the cell network and could include local atmospheric effects, resulting in very high quality GPS coordinates. If one wants to make use of this wealth of geographic data, it is important to educate people about the importance of keeping date information together with location information. This is true not only for the growing number of lay GIS programmers, but may even be relevant for private purposes. For example, if someone wanted to bury a time capsule for their grandchildren, precise GPS coordinates alone would not be sufficient.

The descendents who are trying to find the treasure without a bulldozer also need to know when these coordinates were valid. In order to facilitate explaining these issues to a larger audience, it might be beneficial to use terms like 'reference date' instead of 'epoch' and not to reuse names like 'WGS84' for new 'realizations' that actually change the definition of a spatial reference frame.

3. A Workaround

Assuming one knows the epoch that was used for georeferencing, the following describes a workaround for WGS84 UTM imagery and elevation data. One could use HTDP to find the correct offsets for Northing and Easting for the particular area of interest, define a custom projection that includes this offset, reproject from this custom projection to the usual WGS84 UTM projection, and then import this reprojected "fake WGS84UTM" data into the terrain software (that does not know about the custom projection). Note that this is not strictly a "correct" procedure, because one is simply shifting the projected data in two dimensions rather than shifting the actual three-dimensional geographic data and then projecting it, which would have slightly different results. It would be better if these time based transformations were integrated in standardly used GIS software, so they could be done more easily and with less error.

For individual survey points, this workaround is not necessary. Instead one can use the HTDP software to calculate the correct coordinates at the desired epoch, assuming the data is in one of the datums that HTDP can translate between. Note that this is not always the case, e.g. data is sometimes provided as NAD83 (HARN) or NAD83 (NSRS2007). HTDP has a model for estimating horizontal crustal velocities based on velocity vectors (i.e. the precise direction and speed) for thousands of individual points, additional detail about the faults, and a separate model for estimating the displacements caused by large earthquakes. HTDP can predict velocities with a standard error of less than 2mm/year and a worst case of about 10mm/year at some points right on the San Andreas Fault [10]. But note that some survey data is based on old geodetic markers that are not as precise as a modern GPS survey. For such data, the total error could in fact be much larger than the 1.7m error mentioned above.

If a new survey is conducted, one can instruct the surveyors to do it relative to a two hour GPS occupation, i.e., keeping the GPS operating in exactly the same place for two hours and recording all the data. When correcting the GPS occupation coordinate with the Online Positioning User Service (OPUS), one can then use the ITRF00 (EPOCH:2009.XXXX) output from OPUS to make the correction in the survey software, instead of the usual NAD83 (CORS96) (EPOCH:2002.0000) output. This will propagate the offset to all the other survey points.

This method solves the most serious problem - the position of tracked entities with respect to important buildings in the synthetic terrain. However, it can be hard to explain the procedure to some traditional surveyors, who are used to dealing with NAD83 and may not be sufficiently familiar with these issues, and may lack the right equipment or software expertise. Trying to explain this approach to surveyors several different states has lead to some unexpected results. Due to wrong conversions between UTM and State Plane (Lambert Conformal Conic projection in that state) survey points were displaced in a centrifugal pattern around the central GPS location. The magnitude of the inaccuracy grew with the distance from the GPS point due to a feet to meter conversion mistake. 400 meters away from the GPS point the error was as large as 55m. Even errors of this magnitude can be hard to detect when there are no obvious reference points that are visible in the aerial imagery because the surveyed buildings were only recently placed.

4. Implications for Reusability

The workaround described in section 3 causes the terrain to be in a non-standard datum matching the epoch (date) of a particular exercise. This means that the terrain is not interoperable with other terrains. In addition, depending on the accuracy requirements, the terrain may not be reusable for a future exercise at the same location, either.

Even if a terrain needs to be rebuilt for other reasons, e.g. because there is some new data, it would be helpful if at least the source data that is not new, e.g. previously surveyed buildings, were straightforwardly reusable.

The workaround also adds a complex, semi-manual, and hence time-consuming and error-prone step to an already complicated terrain production procedure.

Another interesting reusability problem involves accuracy across plate boundaries, where the distances and angles between points across the fault lines become less accurate over time. In fact, there is significant deformation for hundreds of kilometers on either side of these faults, affecting all of California and beyond, and making the epoch of the input data much more important there.

5. A Longer Term Solution

Instead of moving the synthetic terrain to match the GPS data at the specific time of each exercise, it would be better if the larger community agreed on a standard epoch for terrains. Then HTDP could be used to transform the live tracking data to match that epoch. But the standardized epoch should not be specific to terrain databases generated by one particular group, to make sure they are interoperable with other terrains.

Alternatively the epoch of the terrain could be specified in its metadata, and HTDP could transform the tracking data to that epoch if the terrain is used at a later time or together with incompatible terrains.

In addition, it is important to raise awareness of the fact that WGS84 coordinates are time dependent, and to make sure that the reference date is always part of the metadata. The exact realizations of both WGS84 and NAD83 also need to be specified in the metadata.

Furthermore, to ensure terrain reusability and interoperability, terrains should also have documented metadata, including not only the spatial reference frame of the completed terrain, but also those of each piece of source data (imagery, elevation, and survey data, including details about the survey methodology), and what transformations were applied to each, by what version of what software. In other words, the terrain provenance metadata specification needs to be recursive and include as much of the data's history as is known, including the epoch of any WGS84 data. This information is necessary in order to correct for software inaccuracies that are discovered later, to calculate the total amount of location difference between terrains, and to determine whether they meet the accuracy requirements of a particular exercise. In order to be machine understandable, this recursive metadata specification should be standardized.

6. Conclusions

The source data for making synthetic terrains is frequently in a different datum (NAD83) than the live GPS tracking data (WGS84). These datums were the same in 1984, but plate tectonic movement and changes in the realizations of these datums have caused them to differ significantly, by about 1.7 meters, depending on the location. These changes are not handled by current software.

The paper suggests both a workaround and a longer-term solution to this problem. The workaround is to use HTDP to calculate the offset and shift the data accordingly. A longer term solution would be to use HTDP to transform the live tracking data to the epoch of the terrain.

It is also important to ensure that the reference date (epoch) is part of the metadata for all WGS84 data, and that terrains have recursive metadata, including the spatial reference not only of the final product but also of the source data items, and details about the processing steps applied to each item.

Acknowledgements

I would like to thank Michael Beebe, Reginald Ford, Hans Haselbach, David Hanz, Mark Johnson, Theresa Keeler, Cameron Kellough, Gerry Shaw, John Shockley, and Ralph Toms for their comments and helpful discussions.

Glossary

CORS Continuously Operating Reference Stations: CORS stations provide reference information about ionospheric and other errors for OPUS corrections.

Datum The datum defines the position of the reference ellipsoid relative to the center of the earth and points on the earth surface, resulting in a definition of the origin and orientation of latitude and longitude lines.

ECEF Earth Centered Earth Fixed

GCS Geodetic/Geographic Coordinate System. A datum, including a reference meridian and an angular unit of measure such as degrees, defines a geodetic coordinate system in which latitude is the angle formed by the normal to the ellipsoid and by the equatorial plane, and longitude is the angle with respect to the reference meridian.

GIS Geographic Information System

GPS Global Positioning System

HARN High Accuracy Reference Network

HTDP Horizontal Time Dependent Positioning

IERS International Earth Rotation and Reference Systems Service

IRM IERS Reference Meridian

ITRF International Terrestrial Reference Frame:

A Realization of ITRS

ITRS International Terrestrial Reference System: A geocentric coordinate system with a condition of no net horizontal tectonic motion

LVC Live Virtual Constructive

NAD North American Datum

NGS National Geodetic Survey

NSRS National Spatial Reference System

OPUS Online Positioning User Service

Reference Ellipsoid A model of the shape of the earth, defined by a semi-major (equatorial) axis and a flattening

RTK Real-Time Kinematic

UTM Universal Transverse Mercator

WAAS Wide Area Augmentation System

WGS World Geodetic System

References

[1] M. Johnson, R. Giuli, S. Oberg, M. Beebe, R. Ford, J. Shockley: "Integration of CCTT and JCATS in an LVC Exercise". 04E-SIW-066, 2004.

[2] M. Beebe, M. Faust, R. Ford, J. Shockley, M. Johnson: "The Creation of GrafenBob for the JTEP LVC Demo". 04E-SIW-069, 2004.

[3] E. Blackwell, M. Moeglein, D. Nakayama: "A global DoD-optimized DGPS for precision-strike", ION GPS-95, pp. 399-411. 1995.

[4] FAA/NSTB: "Wide-Area Augmentation System Performance Analysis Report". Report #17. 2006.

[5] T. Soler, P. Michalak, N. Weston, R. Snay, R. Foote: "Accuracy of OPUS solutions for 1- to 4-h observing sessions". GPS Solutions, 10(1), 45-55, 2006.

[6] R. Snay, T. Soler: "The evolution of NAD83". Professional Surveyor, 20(2), 2000.

[7] R. Snay, T. Soler: "WGS 84 and ITRS". Professional Surveyor, 20(3), 2000.

[8] R. Snay: "Using the HTDP Software to Transform Spatial Coordinates Across Time and Between Reference Frames", Surveying and Land Information Systems, 59(1), pp. 15-25, 1999.

[9] National Geospatial-Intelligence Agency: "Department of Defense World Geodetic System 1984", NIMA Technical Report TR8350.2, Third Edition.

[10] C. Pearson, R. McCaffrey, J. Elliot, R. Snay: "HTDP3.0: Software for Coping with the Coordinate Changes Associated with Crustal Motion". Journal of Surveying Engineering, Forthcoming.

Author Biographies

DR. SUSANNE RIEHEMANN is a Software Development Engineer at SRI International in Menlo Park, CA. She develops various terrain products for the XCTC and JTEP programs at SRI's Geovisualization Center, and also works on terrain ontologies for the ONISTT project.