THE
WYOMING
BUREAU OF LAND MANAGEMENT
FIELD GUIDE FOR EVALUATIVE TESTING OF ARCHAEOLOGICAL SITES

A Supplement to the Wyoming Bureau of Land Management Cultural Resources Handbook

May 2003

BUREAU OF LAND MANAGEMENT
FIELD GUIDE FOR EVALUATIVE TESTING OF ARCHEOLOGICAL SITES

Introduction

Field archeologists should be prepared to test sites where insufficient surface evidence exists to determine the presence of information important to prehistory. Limited subsurface testing may be required for the following purposes:

1. Testing to evaluate significance, usually under Criterion D.
2. Testing to assess the effects of an undertaking on an eligible site.
3. Testing to look for buried cultural remains when there is reason to expect a great potential for significant cultural resources where surface evidence is absent.
4. Testing to develop a data recovery plan for an affected National Register eligible site.

Methods used for all four purposes may be the same or in some situations may differ. It may be possible to conduct two or more kinds of testing at the same time or part of the testing may be useful for two or more goals, while additional testing may be needed to meet a second goal. The archeologist should clearly specify the reasons for testing and which methods and procedures were used.

Poor testing may lead to inaccurate evaluations of eligibility and effect. This in turn may lead to discoveries that result in costly project delays and extra work and expense to mitigate sites that might have been avoided if properly identified. Damaging a site because of an inadequate identification effort is contrary to the intent of historic preservation laws and regulations.

I. Evaluating sites for National Register eligibility.

Evaluating sites for National Register eligibility requires assessing the potential of the site to contain data that contribute to answering important research questions. The following research questions provide examples to stimulate the thinking of the site evaluator. These research questions are only a partial list of possible research values for eligible sites. Historic context (if available) should be referred to for more specific research questions.

Site evaluations must specify what kinds of evidence the site contains and what research questions those data will help answer. Without this information, the archaeologist has not completed a site evaluation. Sites cannot be evaluated without the archeologist interpreting the surface and testing data collected.

A. Research Questions

1. Does the site contribute a better understanding of cultural chronology or a particular cultural phase?

2. Does the site contribute data on subsistence strategies and practices?

3. Does the site contribute data on settlement patterns, landscape use strategies, duration of occupation, seasonal migration round, or mobility strategies?

4. Does the site contribute data on the functions of site structures, such as hearth types and their uses, house types and their uses and duration of use?

5. Does the site contribute data on lithic procurement, processing (reduction strategies), transport, role in settlement strategies and seasonal rounds, selectivity for specific tasks, and/or discard patterns?

6. Does the site contain data on paleoenvironments and adaptations to changing environments?

7. Does the site contribute information on religious concepts or practices?

Additional questions about the nature of the site can help the archeologist assess its potential for contributing data that answer important research questions, such as those above.

This checklist of questions helps the archeologist evaluate the site integrity and assess the nature of the site data. A "yes" answer to any of these questions does not constitute a site evaluation. The investigator must go on to determine what research questions may be answered with these data.

1. Is the site datable? Does the site contain carbon or charcoal or temporally diagnostic artifacts, and/or does it have associated historical records that place the site in time. If datable evidence is not present on the surface of a site, testing may be needed to explore for datable materials.

2. Is the site stratified vertically or horizontally? Does the stratigraphic context retain sufficient integrity in at least one stratum or area to produce information that is discrete to that occupation (note: only one discrete unit is necessary)?

3. Does the site have artifacts that allow interpretation of site function?

4. Does the site have definable activity areas?

5. Are there multiple components that are segregated vertically or horizontally such that relative dating is possible?

6. Does the site have artifacts or features that are unique to the area or time period?

7. Does the site have features? If so, what types and how can they contribute to understanding prehistory?

8. What is the site's environmental location in relation to other sites in the area?

9. What is the site's overall size, complexity, and artifact density, compared to other sites in the area?

10. Does the site contain complete tool kits?

11. Do features have good integrity?

12. Are floral or faunal remains preserved?

B. Site Integrity

The issue of site integrity is complex yet essential for determining site eligibility. Evaluations of National Register eligibility require an assessment of seven criteria of integrity: location, design, setting, materials, workmanship, feeling, and association. Within an archaeological context only some of these criteria may apply, however under the National Register guidelines archeologists need to be aware of all the criteria when considering the complex issue of site integrity.

Specifically, archeologists must become aware the factors that could influence the nature and condition of the cultural deposits. These could be, bioturbation, subsidence, slope wash, wind erosion and other post depositional processes. Knowledge of the local geomorphological conditions is essential in assessing site integrity.

C. Methods of Evaluative Testing

A phased approach progresses from minimal to greater impact on the site. The amount of data recovered generally also increases with greater impact methods, except for backhoe trenching. Phases need not be completed in a specified order. Environmental indicators, such as soil type, geomorphology, and topography, may suggest going immediately to a later 'phase' to maximize the most efficient and cost effective recovery of useful data, depending on professional judgment. Evaluation based on surface evidence, shovel or auger probes and to a limited extent formal excavation units can be carried out under the BLM permit for survey and limited testing. Any effort beyond this must be coordinated with the BLM Field Office archeologist and the project proponent. Testing above that specified in the BLM permit must involve consultation with SHPO.

<u>Method 1: Evaluation from Surface Evidence</u>

Evaluations from surface evidence should use existing exposures of profiles, such as cutbanks, zones of deflation, like blowouts or the windward side of some dunes, naturally occurring disturbances, like animal burrows/backdirt, entrenched animal trails, and ant beds, and man-made disturbances, such as roads. Frequently, an archeologist can evaluate a site based solely on surface evidence if he/she provides sufficient rationale to justify the evaluation. Often, evaluation from surface evidence leads to an ineligible evaluation because it is easier to assume that National Register qualities are absent than to recover incontrovertible evidence of information important in prehistory. Many sites can be evaluated by surface evidence alone, given adequate knowledge of the site and its context. If the existing site conditions do not provide sufficient evidence to answer the basic research questions and information needs, however, some sort of subsurface investigation should be made using one or more or the following methods. The *assumption* that no subsurface deposits are present because the vegetation cover is sparse is not acceptable. Similarly, the assumption that no subsurface deposits are present because of dense vegetation cover, such as in a hay meadow, is not acceptable.

Method 2: Remote Sensing

Archeologists are beginning to use remote sensing techniques for evaluating locating and evaluating sites. As the technique becomes more refined it may enjoy a broader application. Remote sensing is most useful for specific kinds of site types and deposits. For example, if the archeologist thinks the site has a high probability to contain one or more buried hearth(s) or pit house(s), remote sensing would be the least site disturbing technique and may be the quickest way to obtain the evidence.

In terms of site testing and identification, remote sensing techniques may be most useful for preventing discovery situations and searching for buried cultural deposits in contexts lacking surface cultural evidence. Remote sensing techniques like magnetometry, gradiometry, soil resistivity and ground penetrating radar require a research plan which contains strategies for grid coverage and anomaly evaluation and testing. Follow-up construction monitoring to assess the reliability of the method used may be required. See Section III (Testing for buried cultural remains) for a review of remote sensing techniques.

Method 3: Shovel Testing or Auger Probing

This method is characterized by small diameter probes to identify the depth, nature, and extent of cultural deposits, and as an efficient means of determining where further testing may be conducted. This technique uses small-diameter soil cores, augers, or shovels to probe sites, resulting in the least impact while providing information about buried site potential. The method is useful for mapping depth and extent of deposits and serves as an efficient means of determining where further testing efforts will be conducted, if necessary. The method also works well for probing known buried sites to map the extent of buried cultural components for development of data recovery plans.

The degree of success in applying this technique depends on the type of tool employed and the ability of the archeologist to interpret the results. A 3" diameter soil core or larger bucket auger works well, while power screw augers do not. A method which recovers a core that can be analyzed for soil stratigraphy, inspected for depth of artifacts, and screened or troweled to recover the artifacts provides more information that a screw auger that churns up the soil and provides only an artifact count, lacking depth and context information. Shovel tests allow observations about stratigraphy and other aspects of the deposits that auger probes do not always provide. Inspecting larger amounts of soil in shovel tests than in small diameter probes increases the chances that the test will encounter cultural remains that are present in the subsurface. The spacing of the probes, the manner by which augered sediments are examined, and the way positive results of probes are interpreted affect how decisions are made about continued testing based on those results. The spacing and pattern of probes is critical to the viability of the method, as is size of screen mesh used to sift deposits--closer interval probes produce more reliable information and smaller mesh size increases potential for positive recovery of data.

Major drawbacks for shovel and auger probing are the limited depth the archeologist can reach with shovel tests and some types of augers. Testing deeper deposits may require the archeologist to select Method 4 (controlled tests) instead. If the archeologist has a high expectation that the site is eligible, but simply needs evidence of subsurface deposits, he/she should excavate a small-sized controlled test pit because the test will recover significantly more data that a shovel of auger probe. Augering or shovel test probes frequently do not recover sufficient horizontal and vertical data on the remains. Small diameter probes or larger auger or shovel tests on large sites can by chance completely miss the subsurface evidence because the volume of soil inspected compared with the site=s total volume is extremely minimal. The most frequent error in testing is not placing the units in the best possible locations. The locations should be selected based on the site geomorphology, patterns of prehistoric land use, and other factors. The archeologist must justify the rationale chosen for the probe locations.

The type of testing pattern depends on the site. For example, sand sheets need a grid pattern, whereas sand dunes may need probes along the dune topography. Probe spacing must vary with site expectations. Activity areas produced by hearth-tethered hunter-gatherers tend to occupy about a 4 m radius around the hearth. A 10 m interval may not serve the purpose of identifying significant cultural remains, while a 5 m interval will have more likelihood of succeeding. Procedures should be discussed with the BLM archeologist during the fieldwork, to minimize discoveries and return field trips and to maximize confidence in the validity of testing results. Deposits from probes will normally be screened through 1/8 inch mesh in sand and loose soils and where small finds are expected, such as seeds, charcoal, retouch flakes, or small animal bones. A 1/4 inch mesh may be used depending on the characteristics of the deposits, such as heavy or wet clay or in contexts where larger mesh is expected to recover the type of remains the site is likely to contain.

During the testing process, investigators make decisions concerning how to proceed--whether to tighten the testing interval, dig formal test units, or terminate the effort due to sufficient information collected. Decisions on continuing or terminating a testing program should be made with attention to the area of potential effect (APE), the portion of the site containing intact deposits, and the present and future effects, direct and indirect, to that portion of the site.

The combined surface data and site probing results should be interpreted to determine what kinds of archeological evidence are present or expected and what research questions they can help answer. The probing should also be designed to provide information about the site integrity. An evaluation of site integrity is necessary to establishing the probability that the site will contribute to answering important research questions.

Method 4. Controlled Formal Test Units

Formal test units are usually more labor-intensive and impact a greater area of the site, but provide the most reliable and accurate information to answer the research questions and evaluation needs. However, use of formal test units may require a BLM testing permit. The method calls for careful excavation using archeological techniques that provide controlled data on vertical and horizontal reference points so that original locations of artifacts and features can be reconstructed. Formal testing, in which the test units are carefully located, often provides significantly more information about a site than probing. By providing more intensive data, one or a few formal tests can be used instead of a series of probes for evaluating some sites. If a few formal tests can replace a series of probes, the two methods may not vary greatly in labor intensity.

Specific circumstances may necessitate formal testing or make it the most useful method. Soil depth will often necessitate a test unit, because of the limitations of probing tools. If deposits over one meter deep are known or suspected, it may be prudent to start testing with a formal unit. If during probing, a buried component or deposit is suspected but not confirmed, a formal unit is standard procedure. Other surface indications, such as topography, geomorphology, or the presence of features may be reasons to choose formal test units instead of probing. If buried deposits are suspected, a single formal test, such as a 50 x 50 cm or 1 x 1 m unit, may be sufficient to establish the site eligibility and additional testing would be unnecessary. If auger or shovel testing identified evidence of a cultural feature like a hearth, the probe should be expanded to a formal test unit to provide information about the integrity of the old surface from which the feature emanates.

In deposits less than one meter in depth, a small-sized formal unit, such as a 50 x 50 cm or 1 m x 50 cm unit, can efficiently provide more information about a site than a series of probes because data about the horizontal and vertical locations of cultural remains and soil stratigraphy are recovered. These tests also provide data on the volume/density of cultural materials. By testing a larger volume of soil, formal units increase the chances that the test will recover artifacts, features, or contexts that indicate what kinds of data the site contains. For the portion of the site evaluation that determines whether the site contains data that may answer research questions, formal tests are often the best testing methods. Shovel/auger probes often become merely a check for presence or absence of artifacts or an artifact count with no information on depth or location of cultural remains. If the surface data already provide information about the site's potential to contribute data to research questions, then shovel/auger probing may provide all the additional information needed to complete the site evaluation. If surface data and initial probing do not provide adequate data on the nature of the site for answering research questions, then formal testing should be conducted.

Depending upon the size and nature of the site, formal testing may entail enough additional time and expense that the applicant may want a role in the decision to take that step. For most sites, four Aperson-days@ may be needed to accomplish limited formal testing. If more than four person days are expected, the BLM archeologist should be consulted and may decide what further steps should be taken.

Determining how much testing is enough, whether using Method 3 or Method 4, depends upon understanding the local and regional archeological expectations for site structure. A <u>context for site structure</u> is consequently necessary for a reliable prediction of the amount of testing necessary for concluding that a site is eligible or not eligible. The context would need to include (1) estimated site size and shape, (2) expected density of artifacts and features, and (3) the distribution patterns of artifacts and features.

Few syntheses of such data are currently available for Wyoming although comparisons with other regions are helpful. For example, calculations based on Eastern Woodlands

Archaic sites showed a density of 0.23 artifacts per square meter and 0.004 features/square meter for one site and 0.54 artifacts versus 0.005 features/square meter for another site (McMannamon 1984). The expectation of encountering a feature or high density of artifacts for either site in only a few tests would be extremely low. Because of the aggregation of artifacts and features, a hunter-gatherer site would normally have considerable empty space within the site boundary. Thomas (1986) calculated, based on excavation data and ethnoarcheological analogies, that a hunter-gatherer camp occupied by 2-7 nuclear families would have a nuclear area 5-7 meters in diameter surrounded by a peripheral activity zone which is 50% empty space. Yellen's Bushman camps had cultural debris occupying only 19 to 47.6% of the space within the site area. A sampling interval greater than 8-10 meters would generally miss the core area of an occupation and would have a high probability of producing no artifacts or a low density of artifacts.

The factors to be decided in determining how much testing is enough include (1) the volume tested (size of test pits), (2) the sampling interval, and (3) the sampling geometry. Until regional <u>contexts of site structure</u> are developed for Wyoming, the archeologist testing the site must provide a research design giving the rationale for the amount of testing conducted. The archeologist must conduct enough testing to insure that he/she has not examined only the empty space or low density portions of the site. The testing research design should be submitted with the application for a testing permit. In a number of states, this is called a Phase 2 survey.

<u>Method 5. Mechanized Testing</u>

The use of heavy machinery, such as a backhoe, is not a standard technique to evaluate a site for National Register eligibility. However, it may be appropriate for specific sites, particularly those in alluvium or deep eolian or colluvial deposits. Mechanized testing involves considerable site disturbance which could destroy National Register qualities and will be carried out only with an approved backhoe testing and discovery plan. Backhoe trenching is always inappropriate for testing sites with a shallow stratigraphy.

Backhoe testing is most commonly used to investigate site stratigraphy or geomorphology. Backhoe trenches through cultural deposits largely destroy all data within the trench disturbance. Backhoes are useful for finding visually obvious cultural deposits such as major charcoal-stained lenses, hearth features, and bone beds. Less visually apparent remains often cannot be seen on the disturbed walls of the trench. Mechanized testing is consequently appropriate as a testing method largely where time is limited and the site will be impacted. Most BLM projects are not of this nature.

Sometimes, information about site stratigraphy and geomorphology may be important to establishing site eligibility and backhoe trenching may be used to collect or establish the presence of such data. Geoarcheologists and soils scientists may use backhoe trenches in their geomorphic work. The backhoe has useful applications in energy development projects where heavy equipment is easily available. Nonetheless, heavy equipment work on archeological sites presents real hazards to both the site and to the archeologist, so use of a backhoe should be supervised only by qualified archeologists in close coordination with the BLM.

II. Testing sites to assess the effects of an undertaking on an eligible site.

Following a determination of eligibility, it may be necessary to conduct additional testing to evaluate the potential effect of an undertaking on the site. Particularly when the site was determined to be eligible based on evidence outside the APE, it may be necessary to test for effect within the APE. Prior to assessing effects, the investigator must adequately evaluate the entire site. The portion of the site in the APE cannot be considered non-contributing if the site has not been fully evaluated.

<u>Method 1: Evaluation from Surface Evidence</u>

For some sites, surface evidence may be adequate to determine that subsurface deposits are not present or are too disturbed to retain adequate data to contribute to important research questions (see Examples B and C in Appendix A). An assessment based on surface evidence must be adequately justified (see I.C, Method 1).

<u>Method 2: Shovel/Auger Probing</u>

If the site has already been determined to be eligible, limited shovel or auger probing may establish the presence of subsurface deposits of similar nature to those identified when collecting evidence for the site eligibility. A few probes may be all that are necessary to determine an adverse effect on the site area within the APE. Establishing that the APE portion of the site does not contribute to the site's eligibility generally requires more work because a few probes can miss the areas with intact deposits. Probes should be placed where surface evidence indicates the likelihood of intact soil deposits and knowledge of settlement strategies suggests that cultural remains are likely. Placing probes in spots where they are unlikely to encounter cultural remains will not be considered sufficient for establishing the absence of significant deposits.

<u>Method 3: Formal Test Units</u>

In some cases, shovel/auger probes may be inconclusive in establishing the presence or absence of significant cultural remains in the APE. Particularly when probes locate a few artifacts, but do not provide sufficient information about the nature and integrity of the deposits, formal test units should be excavated. Formal test units can provide more information about site integrity because more data on soil stratigraphy and the contexts of artifacts or features are recovered. For example, if the normal soil profile on the site has horizons A1, A2, B2, and C, and the cultural remains are in the A2 horizon, but the tested APE contains only an A, C profile, the investigator may conclude that the site integrity has been destroyed. A single, well-placed formal test unit may be sufficient to establish the presence of significant deposits, whereas one or a few test units can, because of sampling error, miss the intact area(s) of a site. More test units necessary to establish the absence of significant deposits. The interpretation that significant deposits are absent must be adequately justified in the report.

The discussion of how much testing is enough in Section I.C. Method 4 applies also to determining whether a portion of the site is contributing or non-contributing. The archeologist must take into consideration the expected site size, expected artifact and feature density, and expected distribution patterns of artifacts and features. For example, on a project to modify a segment of a road, a 1 x 1 meter test unit produced 3 tools and 2 flakes in a subsurface context. The agency concluded that this portion of an extensive site was non-contributing. The sampling interval was about 100 meters to the nearest test pit. If sites in the region have expected artifact densities of, for example, 0.25 artifacts/square meter, the amount of testing conducted for the project was inadequate to determine that this portion of the site was non-contributing. Additional testing or monitoring of construction should have been conducted.

<u>Method 4: Monitoring</u>

If a reasonable amount of testing has been conducted in the APE, but the results are inconclusive, it may be advisable to stop testing and monitor the construction instead. The decision to monitor should be a part of the determination of effect. Monitoring may take many forms and will usually require an approved monitoring/discovery plan. For large or long linear projects, it may be more cost effective to test site areas within the APE using remote sensing, or another evaluative method prior to monitoring. Adequately evaluating the APE via techniques such as remote sensing can avoid costly discovery situations, which frequently result from poorly planned or conceived monitoring.

III. Testing to check high potential areas for buried cultural remains where surface evidence is absent.

Several geographic contexts in Wyoming are likely to contain buried sites that lack surface manifestations. Failure to identify these areas during inventory may result in unexpected discoveries and result in all the inherent problems with their management. Known geographic contexts include but ar not limited to sand dune fields and extensive eolian deposits, interior basins that lack drainage outlets, alluvium along drainages, colluvium along slopes and scarps, and forest soils under thick vegetation. If you are unfamiliar with the area you will need to contact the BLM Field Office archeologist to determine if any special evaluative treatments are needed. Each context may require a different testing methodology.

<u>Method 1: Remote Sensing</u>

In recent years, remote sensing techniques have been used to aid in evaluative testing efforts. Remote sensing techniques can offer a relatively nondestructive, nonintrusive and highly efficient technology for identifying subsurface cultural materials or features threatened by development. Magnetometer or gradiometer surveys, ground penetrating radar, soil conductivity, resistivity and various forms of infrared and conventional aerial photography have found archeological evaluation applications. Remote sensing usually employs running a sensor over the surface and obtaining readings that report Aanomalies@ or irregularities in subsurface soil characteristics. These anomalies are then interpreted as being of cultural or other origin. This interpretation can, and usually does, take the form of excavation of a test unit, either a shovel probe or a formal unit. Some remote sensing techniques can be sophisticated, require technical expertise and equipment, and are of variable utility, depending on many factors.

Important factors that affect the success or failure of remote sensing applications include soil type, differences in soil strata, nature, depth, density and composition of the target cultural deposit, past disturbances to the surface and subsurface, presence of impurities or contaminants that affect anomaly identification, and other factors. Remote sensing projects frequently employ more than one technique; use of two (or more) systems is currently in vogue, e.g., use of a gradiometer survey in consort with resistivity.

A moderate amount of technical knowledge is needed to assess the utility of such technology. Generally, historic period sites can benefit from magnetometry, gradiometry, resistivity and conductivity. Voids such as are associated with burials, coffins, mine shafts and other open spaces may be detected with ground penetrating radar. Expansive surface features such as trail ruts, ditches roads and linear mounds can be enhanced using infrared or other low or high level vertical aerial photography. Magnetometer and gradiometer surveys are directly affected by certain magetic metals, such as iron. Since historic period sites may or usually do contain metal (and iron is common), the effect of the presence of metal on-site needs to be evaluated. If the location of metal objects is desired, then magnetometry may be eminently suited as a locational technique. However, if scattered historic metal is considered to have contaminated the site, or is so omnipresent on-site (as in some mines, historic debris scatters, dumps, etc.) then the presence of metal may have severe negative effects.

Depth of penetration into the substrate is a critical element in assessing the utility of remote sensing strategies. Most modern gradiometer, magnetometer, resistivity and conductivity instruments can penetrate to a depth of 1 to 1.5 meters only. If the target stratum is buried deeper than this, then remote sensing may be inapplicable. The exception is ground penetrating radar, which can penetrate several meters below surface. Trial and error are frequently needed to properly define a successful remote sensing strategy.

Method 2: Hand Excavation

Locating buried sites that lack surface manifestations by subsurface testing depends on two factors. One is the probability of intersecting a site by a given testing methodology. The testing factors that need to be decided are test pit size, spacing, and layout (Shott 1989). Test pit size has a complex relationship with site density and other factors (Nance and Ball 1989). A rough estimate of the minimum site size likely to be discovered by a particular interval spacing of test pits can be calculated. The formula is: site radius = interval spacing divided by the square root of 2 (Krakker, Shott, and Welch 1983; Lightfoot 1986). To intersect a site of 30 meters diameter or greater, the space between test probes would need to be 21.2 meters (Zeidler 1995). Theoretical and experimental studies of sampling layout indicate greater effectiveness of offset or hexagonal grids compared with square grids (Shott 1989).

Actually finding a site, however, depends also on the probability of detecting a site. Test pits falling in the blank spaces between artifact and feature clusters will fail to detect a site. Both the artifact density and the distribution are key factors affecting detection. As stated earlier, we need to develop regional contexts of site structure in order to design effective subsurface testing strategies.

A. Thick Vegetation Contexts. Where the surface vegetation cover is > 90%, the archeologist should consult the BLM Field Office archeologist about the need for subsurface testing and, if needed, prepare a testing research design. Such contexts include heavy timber with understory vegetation or forest floor litter and dense grasslands, whether natural grasses or cropland, such as hayfields. Archeologists who have conducted experimental and theoretical studies of subsurface testing in dense ground cover contexts are not in agreement on the effectiveness of subsurface testing, particularly in relation to the cost. Sites in forested contexts in Wyoming are usually small in size and consequently are difficult to detect. Subsurface testing to search for unexposed sites in this context should be conducted only in the areas where the greatest disturbances are expected, including logging roads, landings, slash burial pits, and other places where the subsurface will be significantly disturbed. Grassland areas with low visibility are much more limited in areal extent. If the probability of finding a site in these contexts is high, subsurface testing should be conducted to search for buried sites.

For either context, shovel tests are the most common testing method. The interval between probes should take into consideration the expected site diameter and other factors discussed above.

B. Colluvial Deposits. Colluvial deposits along slopes and in draws are one of the least easy to predict contexts. The Carter Kerr-McGee site was on a ridgeslope in a colluvial deposit that ranged up to about 85 cm deep. It contained occupations dating from Goshen, Folsom, Agate Basin/Hell Gap, and Cody times. The Hawken site is in colluvium within an arroyo. Sometimes the unexcavated portion of the bison bone beds are exposed, and at other times, fill completely obscures the deposits. Locally, one portion of a slope may be eroded, while another contains accumulated deposits. Checking soil survey maps for expected soil depth in a survey parcel should be a required first step prior to conducting a survey. More detailed information about local and regional geomorphology is also needed. Effective testing strategies require regional <u>geoarcheology contexts</u>. A geoarcheology study would tell us the depth and nature of colluvial deposits, but not the probability of finding archeological sites in those deposits. For the latter, <u>regional settlement pattern contexts</u> are needed. In particular, we need to know the regional site density and the locations of sites on the landscape. If the site densities in an area are low, the probability of finding a site in a colluvial deposit may also be low. If site densities are high or sites are commonly located where the colluvial deposits are situated, subsurface testing to search for buried sites will be required.

Sites that are buried up to 50 cm below the surface often have some surface exposure if the landform is adequately dissected or eroded, or has blowouts. Where colluvial deposits are greater than 50 cm in depth, the survey archeologist must consult the BLM Field Office archeologist for the need to conduct subsurface testing to search for buried sites. The size of colluvial deposits often is not extensive and testing the deposits would not be cost prohibitive. The archeologist should test the full depth of the Holocene and terminal Pleistocene deposits if possible. For deposits greater than 50 cm in depth, 1 x 1 meter test units are most appropriate. Larger units may be needed for deeper deposits. For deposits greater than 1 meter in depth, other methods besides hand testing may be used, such as mechanized testing or remote sensing.

C. Alluvial Deposits. Alluvial deposits are common along creeks and rivers in Wyoming. Both flowing streams and intermittent drainages contain extensive alluvium. More geomorphological studies are available for helping predict the locations and nature of alluvial deposits than for colluvium, but not enough synthesis is available to design subsurface testing strategies. Regional geoarcheological contexts must also cover alluvial deposits.

Prior to beginning field work, the archeologist should study topographic maps and soil survey maps to determine where alluvial deposits are expected. The probability of archeological sites being located in alluvium is high. Subsurface testing should be conducted in all alluvial deposits which lack exposed manifestations of sites, unless bank or other exposures provide an adequate area of inspection and the width of the alluvial deposits is small.

Because of the depth of most alluvial deposits, hand excavated test units will generally be 1 x 1 meter or larger units. The tests should extend to the full depth of the Holocene and terminal Pleistocene deposits if feasible. The spacing between hand excavation units shall be based on the expected site size and density for the local area. The research design for the testing shall be submitted with the request for a testing permit.

Alternative methods for testing alluvial deposits can also be developed in the research design, such as remote sensing or mechanized testing. Most mechanized testing methods do a poor job of locating buried sites. Backhoe trenches generally obscure all but the most obvious cultural remains. A series of cores is an appropriate method for studying the geomorphology, but usually fails to find archeological sites.

<u>Method 3: Mechanized Testing</u>

The most common form of mechanized testing employs use of a backhoe. The backhoe can be a useful tool for conducting subsurface evaluations in deep soil situations, dunal and eolian contexts, and in areas lacking surface cultural material but where buried sites are suspected. The backhoe is useful when large surface areas are to be impacted such as well pads, plant sites, pipeline complexes. Deposits such as large sand dunes, lee and falling dunes, other eolian deposits and deep alluvial or colluvial soils may be best suited for backhoe testing. An understanding of topography, and how topography affects soil buildup is necessary to effectively place test trenches where they may encounter buried deposits. Shallow soils should be excluded from backhoe testing, as less intrusive methods can probe these deposits.

A Backhoe Testing Plan is necessary and should include a project map depicting the landforms to be tested and where linear trenches will be excavated. The plan must outline methodology for evaluating any cultural materials identified. The plan must address the contingincy that if the objective of the testing is met by excavation of less that the total number of trenches proposed, the effort should be terminated. The plan must also address safety and reclamation measures. Only the area of direct effect, plus a reasonable buffer (25 to 50 ft. adjacent to larger development areas such as well pads) shall be subject to backhoe testing. Backhoe testing of linear projects like pipelines will be limited to the R/W width.

IV. Testing to develop data recovery plans.

Data recovery plans are developed with reference to specific research designs. Testing to develop a data recovery plan requires obtaining specific kinds of additional information. The nature of the site affects what kinds of archeological methods to specify in the plan. Does the site have floral or faunal remains, features, charcoal, lithic workshop debris, pollen, or phytolyths? The depth, thickness, and density of cultural deposits affects the amount of work required. It may be necessary to establish the locations and sizes of intact deposits, places where materials are most concentrated, or number and locations of features in order to determine the size and location of data recovery excavation units.

When approaching this level of testing, most sites should already have had sufficient data documented to evaluate them and assess that the undertaking will have an adverse effect on their National Register qualities. While these previously collected data may be sufficient to reach conclusions about eligibility and effect, they usually are not adequate to develop a specific data recovery plan and research design that state exactly which data will be collected from which areas and how they will answer the research questions. Further testing is usually needed to develop useful plans, and that level of testing may employ any of the methods identified above, depending on the project circumstances and site type(s).

Testing to develop an adequate data recovery plan must consider a number of factors, including:

1. The type(s) of site(s) under investigation;
2. The area(s) of potential effects within the site(s);
3. The research questions identified in the specific research design and relevant historic context;
4. The data needed to answer the questions;
5. The data recovery sample relative to total site area and the APE;
6. The ability of the identified sample to meet the goals of the plan; and
7. The manner by which data are documented in the field, analyzed, and reported in order to provide useful information that meets the goals of the plan.

Testing to develop data recovery plans will build on the information from evaluative testing and focus further efforts on areas where the National Register qualities, or significant information, are documented. Testing may proceed in the same sequence outlined for other testing programs, and terminate when sufficient data are collected. Bearing in mind the high cost of data recovery programs, the investigators should recognize that this effort will probably require much more detailed information acquired from fairly extensive test units in order to justify a valid, well defined data recovery. For example, probes or shovel tests may provide clues about the location and horizontal extent of buried cultural occupations, but controlled test units may be needed to acquire sufficient evidence to justify inclusion of the area in the data recovery plan.

In new sites that are identified through construction monitors or open trench inspections, where no surface evidence was originally identified and little information exists, testing will usually start with hand excavation of controlled test units to evaluate the exposed evidence. If the site is determined to be appropriate for further data recovery, additional testing must be conducted in undisturbed areas adjacent to the discovered evidence in the same sequence outlined for other testing programs. These discovery sites will probably require more data collection at this stage because no other information exists.

V. Site Testing Report Requirements

Inventory reports documenting site testing will contain the following information:

1. If site was not tested, state why testing was not necessary.

2. Site map showing location of tests in relation to exposed features, artifact concentrations, tools, datum, etc.; positive and negative test results must be indicated on the map.

3. A discussion of the testing strategy used with emphasis on why that particular stategy was employed.

4. A specific discussion of the results. This includes:

   A. Description of soil profiles from test units;

   B. Depth and thickness of sedimentary strata encountered;

   C. Sizes and depths of test units and why they were terminated;

   D. Total area tested and percentage of site tested;

   E. Illustrated profiles and plan views from formal excavation units, and profiles of backhoe trenches;

   F. Plan views and profiles of exposed cultural features (if applicable);

   G. Photographs of excavated and profiled features, test units and trenches;

   H. Discussion of special samples recovered and results of their analyses (radiocarbon, flotation, etc.);

   I. Discussion of the relationship of the soil profile and cultural material;

   J. Discussion and full report of the cultural materials recovered: their depth, densities and locations in the test units; the types of tools and debitage and their materials; bone identifications and descriptions; and description of any other evidence relevant to the cultural occupation (raw data may be presented in summary tabular format).

5. Conclusions and recommendations about site eligibility, effects and further work.

REFERENCES CITED

Krakker, J. J., M. J. Shott, and P. D. Welch
1983 Design and Evaluation of Shovel-test Sampling in Regional Archaeological Survey. Journal of Field Archaeology 10:469-480.

Lightfoot, K. G.
1986 Regional Surveys in the Eastern United States. American Antiquity 51:484-504.

McManamon, F. P.
1984 Discovering Sites Unseen. In Advances in Archaeological Method and Theory, Volume 7, edited by M. B. Schiffer, pp. 223-292. Academic Press, New York.

Nance, J. D., and B. F. Ball
1989 A Shot in the Dark: Shott's Comments on Nance and Ball. American Antiquity 54:405-412.

Shott, M. J.
1989 Shovel-Test Sampling in Archaeological Survey: Comments on Nance and Ball, and Lightfoot. American Antiquity 54:396-404.

Thomas, P. A.
1986 Discerning Some Spatial Characteristics of Small, Short-Term, Single Occupation Sites: Implications for New England Archaeology. Man in the Northeast 31:99-121.

Zeidler, J. A.
1995 Archaeological Inventory Survey Standards and Cost-estimation Guidelines for the Department of Defense. Volume 1: Main Text. US Army Corps of Engineers, USACERL Special Report 96/40.