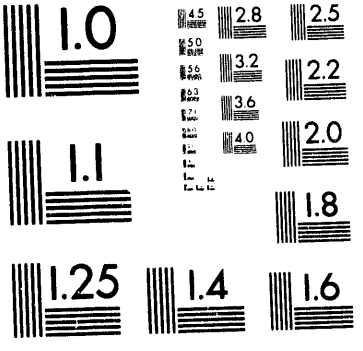

1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
7.1

1 of 1

Conf-9309139--8

LA-UR- 93-2555

*Title:* A CONTINUOUS FLOW FROM SAMPLE COLLECTION TO DATA ACCEPTABILITY DETERMINATION USING AN AUTOMATED SYSTEM

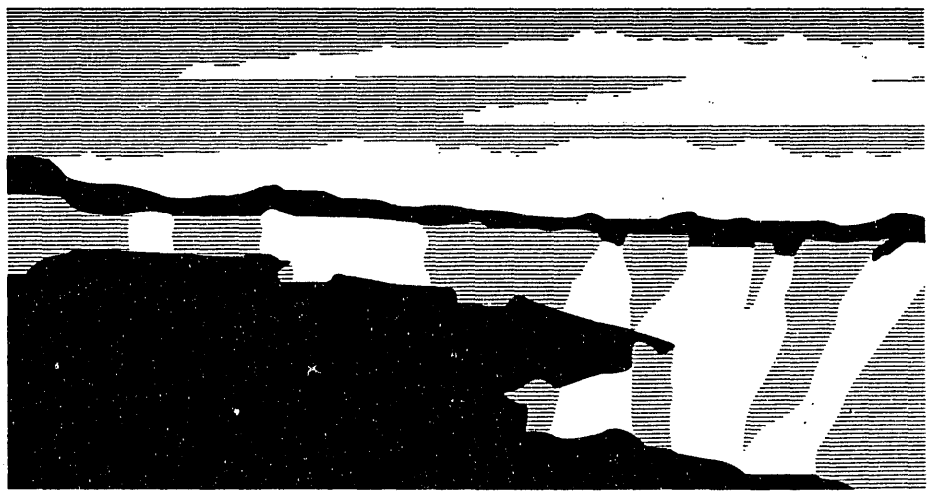
*Author(s):* J. F. Fisk, and C. Leasure



*Submitted to:* Amer. Society for Quality Control (ASQC)
Twentieth Annual Nat'l Energy & Environmental
Quality Div. Conf.
Indianwells, CA
September 19-22, 1993







Los Alamos
NATIONAL LABORATORY

TWENTIETH ANNUAL NATIONAL ENERGY
AND ENVIRONMENTAL QUALITY
DIVISION CONFERENCE

# DATA MANAGEMENT SYSTEMS IN PERFORMANCE MEASUREMENT

## A CONTINUOUS FLOW FROM SAMPLE COLLECTION TO DATA ACCEPTABILITY DETERMINATION USING AN AUTOMATED SYSTEM

Joan F. Fisk
Staff Member and Technical Team Leader
for Environmental Chemistry
(On Intergovernmental Personnel Act assignment from the
Environmental Protection Agency's Office of Emergency
and Remedial Response)

Craig Leasure
Group Leader, Environmental Chemistry Group

Los Alamos National Laboratory
PO Box 1663, K-484
Los Alamos, New Mexico 87545

Andrew D. Sauter, Vice-president
PAR Enterprises, Inc.
217 Garfield
Las Vegas, Nevada 89014

## ABSTRACT

Los Alamos National Laboratory (LANL)is in the process of implementing an automated system that will be used from the planning stage of sample collection to the production of a project-specific report on *analytical data quality*.

Included are electronic scheduling and tracking of samples, data entry, checking and transmission, data assessment and qualification for use, and report generation that will tie the analytical data quality back to the performance criteria defined prior to sample collection.

Industry standard products will be used (e.g., ORACLE, Microsoft Excel) to ensure support for users, prevent dependence on proprietary software, and to protect LANL's investment for the future.

# TWENTIETH ANNUAL NATIONAL ENERGY AND ENVIRONMENTAL QUALITY DIVISION CONFERENCE

## BACKGROUND

The Environmental Protection Agency (EPA) was established on December 2, 1970. This was followed by twenty years of environmental legislations being enacted and then reauthorized, generally with increased authorities. In this twenty year period, environmental data collection activity has constantly increased as a result of such legislations as the Safe Drinking Water Act, the Clean Water Act, the Comprehensive Environmental Response, Compensation, and Liability Act (Superfund) and its second generation Superfund Amendments Reauthorization Act (SARA), the Clean Air Act, and the Resource Conservation and Recovery Act (RCRA).

A huge marketplace has emerged where sampling and analyses has become big business. Whenever "*business*" becomes a watchword, *competition* is on its heals. Competition naturally leads to improvements in technology, to do things better and faster. In the environmental analytical community this has meant *automation* for both the instrument manufacturers anxious to capture laboratory business and for the analytical laboratories trying to capture Government business and related private sector business.

In the earlier days of the new national focus on cleaning up an environment polluted by generations of naive abuse, there was little in the way of automation in the analytical laboratories; in fact there was little in the way of an analytical community prepared to handle large volume routine analyses of environmental samples.

EPA published analytical procedures in December 1979 for analyses of environmental samples (water, at the time) for organic and inorganic constituents in support of the Clean Water Act. Those methods formed the basis for the commonly-used techniques still used today (gas chromatography - GC, gas chromatography/mass spectrometry - GC/MS, inductively coupled plasma - ICP, and atomic absorption - AA). As the analytical laboratory community developed in response to the needs created by the legislations, the age of the microcomputer was also, fortunately, coming to the fore, and was given great legitimacy by the instrument manufacturers and the laboratories as a way to increase productivity. The first evidence of automation seen by Government clients was computerization of reporting forms - quite an improvement over the handwritten forms on which data had been reported. Of course, this automation brought with it a new set of problems, including such things as error introduced in manual entry of data, error introduced by computers "talking" to each other, and the ability to alter data electronically (files or forms) to appear to meet a specified requirement when it was not met (several environmental laboratories/personnel were indicted and convicted on various counts of fraud over the past five years).

Despite new problems that occurred with increasing automation, the benefits far outweigh the problems. EPA, in its efforts to increase productivity on its end, as the recipient of analytical data, has taken advantage of the same strides in computerization as the manufacturers of instruments and the laboratories. In its role as regulator or overseer of the regulated community, EPA is the recipient of enormous reams of analytical data - especially within the Superfund Program. In order to better manage the large volume of paper that comes in daily from the laboratory community, Superfund has, for several years, required its laboratories to provide data that is contained on reporting forms to be delivered also on a diskette for uploading to the LIMS for entry into data bases for various purposes such as checking for contractual compliance, tracking quality assurance parameters, and ultimately, for reviewing the data by computer. This last area, automated review of the data, is becoming a reality for EPA personnel for Superfund data generated by the Contract Laboratory Program (CLP), at least for review of organics analysis data. However, these automated programs are not necessarily appropriate for use by clients other than Superfund or those using the exact specifications in Superfund's contracts that are used by the CLP.

Such is the case with Los Alamos National Laboratory's Environmental Chemistry Group (LANL/EM-9) and its emerging subcontractor laboratory community, designed to meet today's needs of the Environmental Restoration (ER) Program at LANL and to take advantage of the state of development and sophistication of both the analytical laboratory community and the constantly-evolving computer technology that can be adapted across the board in almost all facets of an environmental data collection activity.

## "THE PROBLEM"

The same problem that EPA has faced, and has been dealing with for many years, is now a reality across Government Agencies involved in environmental data collection in response to EPA's various legislations - all of which Federal Facilities must comply with and whose standards must be met during clean-up/restoration. The problem has become quite well-defined as a result of meetings held this past year by the Interagency Steering Committee for Quality Assurance for Environmental Measurements[1]

**The problem, in simplistic terms, is a data management problem. It is characterized by the following elements:**

- **Large paper volumes of analytical data often accompanied by an electronic deliverable that is likely inconsistent in**

---

[1] "Highlights of the Workshop for Development of an Environmental Life Cycle Data Base,", Fisk, J.F., June 1993.

**format and contents from any other program's electronic deliverable.**

- **No quick and easy way to review the data for correctness, completeness, and quality.**
- **No way to transmit the results and associated metadata across Agencies, let alone across Programs within Agencies.**

Most recently, (May 4-6, 1993) a 3 day workshop was held by the Steering Committee whose goal was stated to better define what is needed to develop and implement an Interagency system that would (1) shorten the data assessment process, (2) identify a standard data input, (3) replace program-specific requirements with an Interagency system, and (4) provide a flow from instrument level through data review/assessment to the user. As the meeting progressed, details of what an Interagency system needed became evident and include: a database(s) (which should not be a "data dump" for all data), data dictionary, definition of units, standard transmission protocol, common access language to accommodate multiple databases, change control, linkage between Data Quality Objectives (DQOs) and data generated, and maintenance of a base of required information (metadata - to be defined at the 7th Annual Workshop of the Steering Committee for Quality Assurance for Environmental Measurements).

LANL, being one of the Department of Energy's weapons complex facilities that is involved in its Environmental Restoration Program, is well aware of the problems described above and is actively participating in the solutions, on an Interagency basis. However, recognizing that an Interagency system may require several years to clearly define, develop, test, and implement, LANL is taking advantage of the present Interagency thinking in its specific efforts to implement an automated system, wherever possible, for its data collection activities.

## LANL/EM-9 PERSPECTIVE

LANL/EM-9 has decided to implement an electronic system that incorporates the entire life cycle of a data collection activity as it relates to the areas in which EM-9 is involved. In implementing this decision, EM-9 has chosen to use an approach that relies on industry standard software products in order to (1) assure continuing support for both LANL/EM-9 and their community of sub-contractor analytical laboratories, without having to get tied up with proprietary software products and (2) allow for the flexibility to adapt to new needs that may occur in the future (thus protect the investment).

In addition, LANL/EM-9 has planned its system to perform simultaneous review of an electronic deliverable for compliance with specified requirements (subcontract provisions), data quality - as measured against a specified set of performance

criteria, and integrity. *Integrity* review, in this case, can mean either of two things. The first is a review to determine if a laboratory performed manual editing inappropriately to appear to meet a requirement when it did not. The second is a review/check to make sure that the result is correct based on what was done to the sample throughout the analytical process (a check more on "good automated Laboratory practices) and no error was introduced, inadvertently, by anyone or by a computer.

An important feature that is getting as close to a major goal of the Interagency Steering Committee as possible, is the inherent ability to incorporate a set of analytical performance criteria into the LIMS system that may be project-specific rather than a subcontract requirement. The data assessment could be performed using these project-specific criteria as the measure for the analytical data quality. A project-specific report could be produced that relates the analytical data quality to that stated as needed by the client's specifications.

In considering design of the system, LANL/EM-9 was concerned with ease of entry of data by the analytical laboratory and also ease of upload (seamless)into its ORACLE system (LIMS), where data will be stored (likely in a temporary data base) prior to and during review, and will be archived ultimately. It was considered important that there would be ease of download to a PC for data assessment purposes where the data assessors could impart various functionalities for focussed data analysis.

Needless-to-say, the initial primary objective is to perform an automated review of the data in order to provide timely reports to the clients and to be able to focus resources where they have the most value. LANL/EM-9 considers that the "value added" rests with providing useful information to the client about the data quality and how it relates to the client's defined analytical needs, rather than in checking and re-checking what the subcontractor has already checked (and is being paid to check). Therefore, an important facet of the LANL/EM-9 design is the use of built in equations, protected fields, and macros to ensure that the deliverable arrives as error free as possible.

## DESCRIPTION OF "THE INITIAL PRODUCT"

LANL/EM-9 is using Microsoft Excel spreadsheets for both the ordering of the samples and for data entry and delivery to LANL/EM-9. The spreadsheets are password-protected, encrypted (compliant with specifications of EPA's "Good Automated Laboratory Practices Guidance), and contain the data for an entire set of samples. When printed out, the spreadsheets represent about 20% of the paper that is usually generated as reporting forms for typical analytical data such as that required for EPA's CLP. In addition, processed data will also accompany the spreadsheets in electronic format as imported ASCII files . This is important when recreating results when *data integrity* is

an issue being addressed during review.

The use of Microsoft Excel has allowed for the use of pull-down windows that provide the means to evaluate specific QC parameters as they relate to the data and to provide the linkages needed to associate the QC or chain-of-custody information (e.g., instrument identification) to given sample data.

The following section provides a description of EM-9's involvement concerning the data collection activity and relates where electronic communication and assessment tools are being used (or will or can be in the future).

See Figures 1 and 2 which follow the narrative text for a visual picture of EM-9's infrastructure as it relates to data collection. Figure 1 portrays EM-9's involvement prior to sampling and shows the areas of communication that are electronic. Figure 2 portrays EM-9's involvement following sampling up to the generation of a report to the user, or maybe, a group of users who may need the analytical laboratory data for different purposes.

LANL/EM-9 DATA COLLECTION INFRASTRUCTURE

There are two distinct phases to EM-9 involvement in the data collection process - the first being prior to sampling and the second being after samples are collected and then sent to the analytical laboratory for analysis. Note that the analytical laboratory could be a laboratory internal to LANL/EM-9 or it could be a subcontractor laboratory.

**Prior to sampling (Figure 1) the activities include:**

- Discussion between the "Operable Unit Project Leader (OUPL) and/or the OUPL's technical team and EM-9's Data Assessment Team (which also performs a customer service function for ER to help in planning the appropriate analytical services to meet needs as defined during the DQO process) to discuss the nature of the analytical needs, to determine what the performance criteria for the data are, and to establish what "tier" of review is needed for the data when they are delivered. (Note that there are various "tiers" of review based on the use of the data, recent performance information based on objective measures such as QC sample results, etc., and customer demands.) In looking at Figure 1, this interaction is represented by numbers 1 and 2. Number 1 is the client providing project information, analytical data performance criteria, and direction on the needed tier of review. Number 2 is the guidance provided by the Data Assessment Team as related to appropriate analytical services, information on trade-offs concerning tiers of review, and other technical guidance needed by the client. **This interaction is a qualitative process that is objective. There may be written guidance, but no electronic element.**

Numbers 3 and 4 depict the exchange of information between the client and LANL/EM-9's Sample Management function. This notification that samples will be taken must be at least 4 weeks prior to sample collection, as the subcontractor laboratories must be notified of sample delivery at least 4 weeks in advance or they have the right of refusal. Number 3 denotes the notification ahead of time and number 4 represents the acknowledgement of the notice, the assigning of sample numbers, and the advance notice of the scheduled laboratory. **Numbers 3 and 4 will ultimately be performed electronically.**

- Numbers 5 and 6 show the advance notification of sample delivery to the laboratory and the laboratory acknowledgement of the booking. **These are electronic.**

- Numbers 9 shows the entry into the LIMS of the projected sampling activity and number 10 provides information to Sample Management on future loading. **These are electronic.**

- Number 12 shows the entry of the agreed-upon performance criteria and planned tier of review. Number 11 shows information being extracted from the LIMS on projected sampling and analysis so that the Data Assessment Team can plan and manage the data review and laboratory oversight activities. **These are electronic.**

- Numbers 7 and 8 show an interaction between the analytical laboratory selected for the analyses and the Data Assessment Team related to the project, and includes the response to questions that the laboratory may have concerning the project. **This is not electronic.**

**After sampling (Figure 2) the activities include:**

- Number 3 represents the shipment of samples from the field to the Sample Management function. This could also be a direct shipment to the analytical laboratory from the field with electronic notice to Sample Management that shipment has occurred, which would encompass number 5 also. Number 5 represent the shipment to the laboratory, whether from the field or from the Sample Management function. Obviously, this shipment of samples is not electronic. However, number 4 is the electronic notification that samples have been shipped and 5 is the notification that the samples have been received by the laboratory, **both of which are electronic.**

- During the period of time that the samples are undergoing analyses in the laboratory there may be questions related to the analyses (technical problems) which will be addressed as number 8, and guidance provided by the Data Assessment Team is number 7. There may also be questions related to problems with the samples themselves (wrong numbers for assigned samples, wrong matrices, etc.) that are handled in

the interacting shown by numbers 5 and 6. **These are likely all not electronic exchanges**.

Number 9 shows the information being transmitted to the LIMS on the actual sampling, where a tracking software will provide information to the Data Assessment Team as to the status of the samples/analyses so that planned schedules can be confirmed or changed, as necessary. **These are electronic**.

- After the samples have been analyzed, the data reduced and entered onto the electronic spreadsheets, and transmitted to LANL/EM-9 Sample Management (number 6 again), the spreadsheets are uploaded into the LIMS (number 9) into a temporary data base for processing. Then number 11 comes into play when the Data Assessment Team downloads the spreadsheets onto PCs for data assessment and report generation. **These are electronic**.

- The Data Assessment Team performs the assessment process, always doing the electronic review (though sometimes supplemented with hardcopy checking of both forms and raw data). The "validated" results are fed back into the LIMS (number 12) **electronically** and a report is generated that is relevant to the client's needs. The report is then transmitted to the client **electronically.** At this point the loop is closed with the client.

The next step is transmission of results plus specified metadata to "FIMAD," the ER Program data base which contains site/OU information that will ultimately be used to determine if project DQOs have been met, as determined by the Data Quality Assessment (DQA) process. DQA takes into account other sources of bias and variability, such as inherent population variability, survey sampling error, design implementation error and sample acquisition error[2]. LANL EM-9 plans to provide a measure of the error associated with the actual analyses of the samples as an important input into the DQA process. Perhaps, in the future, the goal of the Interagency Steering Committee project of closing the loop on DQOs electronically will be a reality at LANL.

## CONCLUSION

LANL/EM-9 expects to have a system for rapid evaluation of data for compliance, quality, and integrity. It will be able to generate project-specific reports to users that close the loop on their analytical performance criteria and provide needed information for the DQA process. As much of the entire process as possible is automated, or will be in the future, from the beginning to the end of the analytical data collection process.

---

[2] "Data Quality Assessment: Have You Met Your DQOs?", Michael, D., Twentieth Annual National Energy and Environmental Quality Division Conference.

FIGURE 1. PRIOR TO SAMPLING

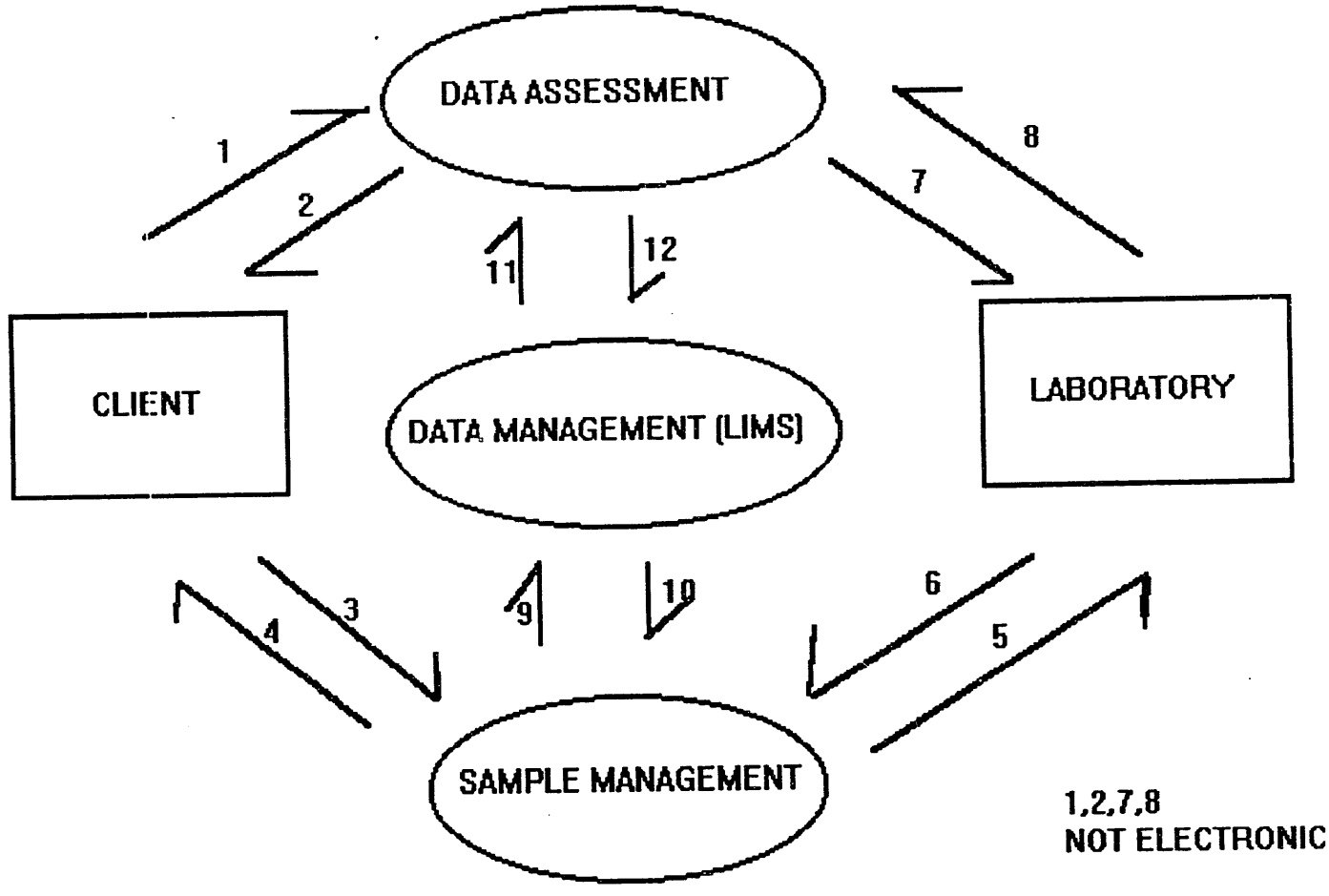
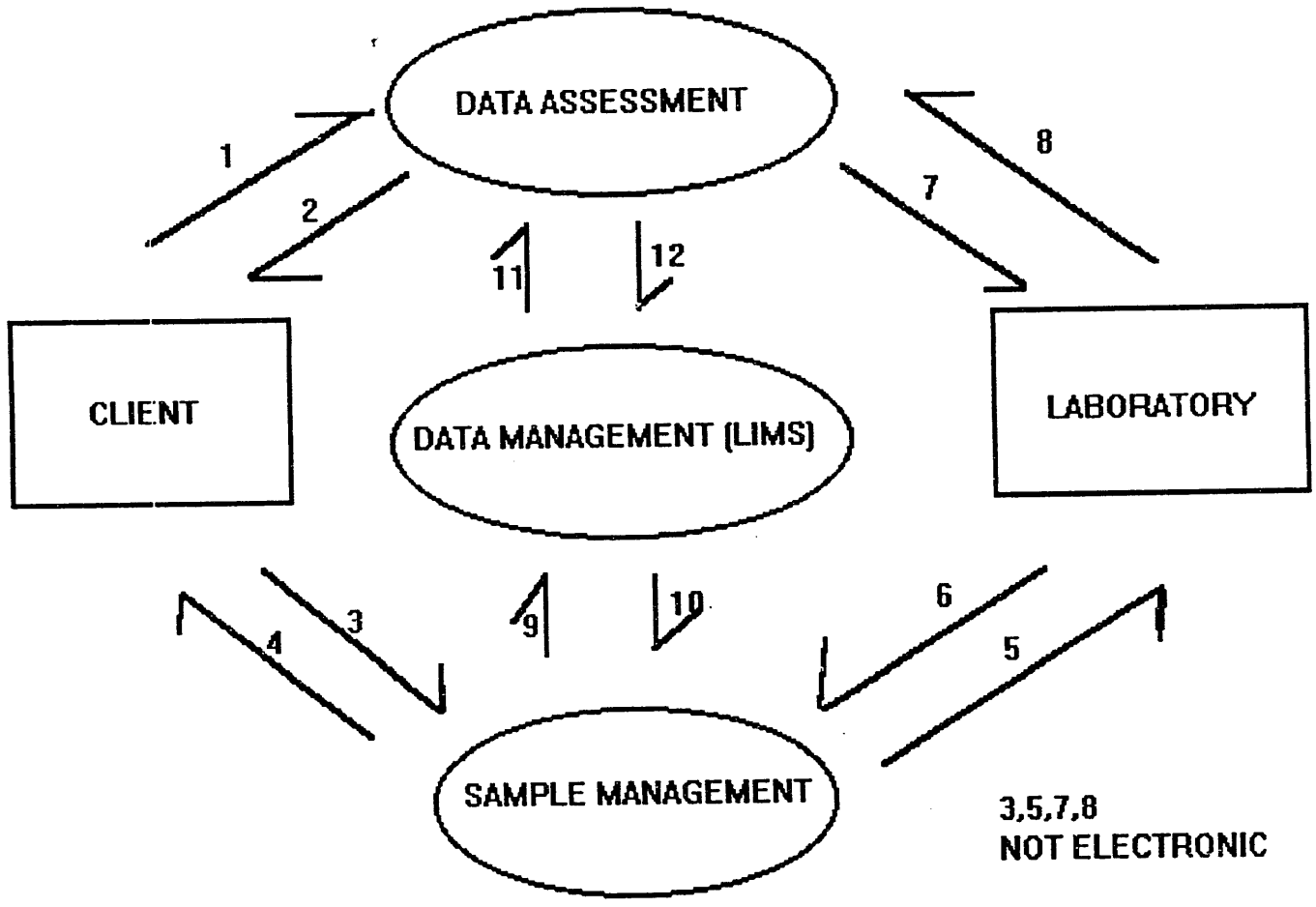


FIGURE 2

AFTER SAMPLING



DATE

FILMED

10/19/93

END