Robert R. Loux
State of Nevada
Agency for Nuclear Projects
Dated: 9/10/07
Enclosures

To: Weber, NMSS
Ref. G20070642
LTR-07-0625

Cys: EDO
DEDMRS
DEDR
DEDIA
AO
RIV

From: CN=Palmer Vaughn/OU=YM/O=RWDOE
PostedDate: 02/23/2007 12:00:35 PM
SendTo: CN=Srikanta Mishra/OU=YM/O=RWDOE@CRWMS
CopyTo:
ReplyTo:
BlindCopyTo:
Subject: Re: Fw: PUT Review Fact Sheet for UZ Parameters & Weighting Factors
Body:

Srikanta, thanks for your professionalism and patience during the discussion yesterday and others such discussions. I appreciate your observations below. I will add the moisture data comparison to the actions. It would be useful to consider your other thoughts below as possible approaches for post submittal activities, either license Defense or Next Generation PA. Palmer

Srikanta Mishra
02/23/2007 07:21 AM

To: Palmer Vaughn/YM/RWDOE@CRWMS
cc: Robert Baca/YM/RWDOE@CRWMS
Subject: Re: Fw: PUT Review Fact Sheet for UZ Parameters & Weighting Factors
LSN: Relevant - Privileged
User Filed as: Excl/AdminMgmt-14-4/QA:N/A

Palmer --

Thanks for your adroit moderation of yesterday's debate. The only suggestion I have to your excellent summary is that we should also look at graphical comparisons between measured and simulated values for the mositure data used in the actual calibration.

Another question that comes to mind is this - if the ultimate objcetive of the UZ flow model is to develop flow fields for transport modeling, then why not use a joint inversion approach that combines moisture data, chloride data and temperature data? Once could then directly use the goodness of fit for each conceptual model (flux map) to derive weighting factors. In fact, this is the approach advocated by Shlomo Neuman and his co-wrokers in the MLBMA (maximum likelhood Bayesian model averaging) methodology.

I think the calibration process is also being constrained by the fact that we impose spatially homogeneous properties for the entire model domain (on a layer-by-layer basis) as opposed to a distributed parameter field that might provide greater spatial resolution (and hence, potentially better agreement between simulated and measured values).

These are all more fundamental questions than the immediate problem on weighting factors, but just my 2c.

-- Srikanta
Palmer Vaughn
02/22/2007 02:18 PM

To: Hui-Hai Liu/YM/RWDOE@CRWMS, Srikanta Mishra/YM/RWDOE@CRWMS, Robert Baca/YM/RWDOE@CRWMS, Clifford Ho/YM/RWDOE@CRWMS, Stephanie Kuzio/YM/RWDOE@CRWMS
cc: Bob MacKinnon/YM/RWDOE@CRWMS, David Sevougian/YM/RWDOE@CRWMS, James Blink/YM/RWDOE@CRWMS, Bill Arnold/YM/RWDOE@CRWMS, Laura Price/YM/RWDOE@CRWMS, laura.l.price@cox.net, Lorenzo Salgado/YM/RWDOE@CRWMS, Robert Baca/YM/RWDOE@CRWMS, Clifford Hansen/YM/RWDOE@CRWMS, Ming Zhu/YM/RWDOE@CRWMS
Subject: Re: Fw: PUT Review Fact Sheet for UZ Parameters & Weighting Factors
LSN: Relevant - Privileged
User Filed as: Excl/AdminMgmt-14-4/QA:N/A

Thanks to all for a very professional debate on this difficult issue: the PUT team for raising their valid concerns, the NS team for justifying their

Template: SECY-017

E-RIDS: SECY-01

approach and taking on some action, and the PASIT for asking tough questions. Attached is my condensation of the issues and path forward. There are some actions, which I assigned dates to. Please let me know if you have concerns over the action items or if I failed to capture something correctly or at all.
Thanks, Palmer
Message Addressees

To:
Srikanta Mishra/YM/RWDOE@CRWMS

Attachment: PUT Review of UZ Flow Weighting Factors.DOC

Corrosion Issues for PASIT Meeting Tomorrow 10:00am to 12:00

1) LC of WPOB
    a. Currently adopting single line fit to all data
    b. Status two line fit
        i. Issue of Exclusion of data
    c. Risk information from TSPA

2) Drip Shield Corrosion for Compliance Model
    a. Corrosion Team to develop
    b. Status

3) GC for WPOB
    a. Current Use 5 year data
        i. Problem 5 year data more realistic
        ii. Results in lower rates
        iii. However, narrower uncertainty results in higher peak
        iv. Contrary to PMA purpose of reducing conservatism
    b. Proposed: Remove from PMA and put in compliance
    c. Corrosion team to approve or provide alternative for PMA

4) Drip Shield Corrosion for PMA
    a. Current Use 2.5 year data for PMA
    b. Does this yield same difficulty as in item 3.

5) Treatment of MIC
    a. Current Screen out on the basis that individual colonies are not present long enough to result in significant localized damage.
    b. General Corrosion will be enhanced as done in the past.
    c. Status form corrosion team

6) LC for Alloy 22 in low Cl environments

    a. Moot issues since condensation under the drip shield is to be screened out

7) Temperature Dependence of GC

    a. Currently using temperature dependence.

    b. Options for TSPA

        i. Continue to use full temperature dependence

        ii. Use Temperature dependence above 60 but not below

        iii. Do not use Temperature dependence.

8) Re-passivation Model

a. Currently not on base line

b. Not doable for compliance or PMA

c. Possibly for consideration in Next Generation or License defense.

WELCOME to the Near-Field Environment (NFE) Group of the OCRWM Lead Laboratory for Repository Systems. On Tuesday, October 3 an all hands meeting was held in Las Vegas to present some Lead Lab kickoff information to Las Vegas-based employees. This e-mail summarizes that general information, primarily for the benefit of non-Las Vegas-based personnel, and provides some kickoff information regarding the NFE Group. In the next week or two I hope to be able to visit in person with all of the non-Las Vegas personnel.

Our mandate as a part of the Lead Lab organization is to produce a credible (i.e., technically competent) and defensible (i.e., compliant with 10 CFR 63 and traceable) License Application on or before June 30, 2008 (i.e., on schedule). This translates to a NFE scope of work that consists of (a) producing AMRs, (b) interfacing with and providing feeds to TSPA, and (c) supporting SAR Section development. The priorities, as emphasized by OCRWM Director Ward Sproat during his visit to SNL on September 21-22, are as follows:

Schedule – If we do not meet the June 30 deadline, "we are all out of a job". Therefore, the short term focus for NFE is on the AMR schedules. All AMRs that feed TSPA must meet the following milestone dates:

December 15, 2006 – Form and Function to TSPA. This coincides with completion of the draft revision of the AMR (check copy).

March 29, 2007 – Preliminary DTN to TSPA. This coincides with completion of the checking phase of the AMR.

May 31, 2007 – Final DTN to TSPA. This coincides with completion and approval of the AMR.

Any slips in schedule will be recovered by cutting scope. There is no allowance for not meeting schedule.

Regulatory Compliance – All of our AMRs will be traceable. Given a choice between a complex, state-of-the-art technical approach that is difficult to validate and/or defend and a simple approach that can be validated and defended with greater traceability. The simpler approach is preferred.

"There may be holes in our workscope, but there cannot be holes in our QA"

Technical Competency – The technical basis will be consistent with the Annual Work Plans (AWPs) as supported by the TWPs. No work, not already identified in TWPs, is necessary. Our mandate is to produce competent, defensible, and traceable work, not unnecessarily complex state-of-the-art analyses. Where greater complexity causes schedule slips, scope will be cut and/or alternative approaches will be considered.

Where greater complexity becomes necessary (i.e., if NRC thinks there are holes in our workscope), it can be introduced during License Defense.

Attached is an organizational chart for NFE. This organization is responsible for 20 Work Packages (i.e., about 20 AMRs), several TSPA feeds, and 3 SAR Sections. This org chart shows all personnel expected to contribute, at various locations, some full-time, some part-time. Those of you currently working on AMRs know who you are, others, who may not be contributing to an NFE AMR right now, will be contacted in the next few days to discuss your role (e.g., assistant author, checker, etc.).

My responsibility, as NFE Manager, is to ensure that the 3 priorities – schedule, defensibility, credibility – in that order, are satisfied. I will get involved in a technical sense only enough to be able to balance those priorities. Given the past successes of the NFE workforce, my main objective is simply to not ruin a good thing.

Detailed technical direction, integration, and decisions about necessary work scope will be provided by the two Technical Leads, Pat Brady (Chemistry) and Ernie Hardin (Thermal Hydrology). The technical areas in parentheses are just for general guidance, there will be several areas/AMRs where they overlap.

Administrative Support (listed as TBD on the org chart) will be provided by Arlene Nery and Patti Weigand.

ACTION ITEMS FOR EACH OF YOU

- Take Annual GET Training (through BSC) by October 31. This training introduces 8 new Lead Lab science procedures (SCI-PRO-xxx).
- Read the 8 SCI-PRO procedures (available on CDIS). They are not significantly changed from the previous BSC procedures.
- Use Lotus Notes for e-mail communication wherever possible. This helps to satisfy LSN requirements. If you must use a different system, you must cc "Lead_Lab@notes.ymp.gov.

That's all for now. Looking forward to working with you all

Geoff Freeze

OCRWM	Comment Sheet	1. QA: NA 2. Page 0 of 17

3. Document Title:	
Yucca Mountain Repository License Application Conceptual Design Report, Postclosure Analysis and Activities **NNPP Comments on SANDIA's comments provided in the "Response" column**	
4. Document No. /Rev./ICN: XXX-XXX-XX-00000X REV 00A	5. Date: July 2006
6. Manager of Reviewing Organization (Print Name): Tito Bonano	7. Org./Discipline: Lead Laboratory Licensing Department

8. CODE	9. SECT./PAR	10. COMMENT/SUGGESTED RESOLUTION	11. RESPONSE (NNPP comments on SANDIA's comments)
1. PA (TSPA)	General	Lists of all tables and figures in the September 2005 version of the LA are provided for each section. These lists take up a lot of space for minimal added value. Consideration should be given to eliminating them. The attempt to identify all of the tables and figures that may change is a little more helpful, but still is probably not the optimum way of presenting this information. It would be much more informative to replace these extensive lists with text that describes the nature of the changes and the reasons for them. We probably cannot predict precisely which specific figures will change, nor should this be prescribed. Some flexibility in what figures and tables will best explain the results associated with changes presented in the SAR is appropriate and necessary.	Agree.
2. PA (TSPA)	General	Most of the lists of tables and figures anticipated to change provide no reason for the changes except something like "May change due to changes in _____." It would be helpful to list the specific reasons for the changes if they are known.	Agree. Need to evaluate Rev 0C for continued applicability of this comment—additional information added to Rev 0C may negate its applicability. (AS)
3.	General	Risks are presented in several different tabular formats. The format should be standardized and used throughout the document. The most helpful format is the one used in Section 12.6 on page 226, as it presents estimates of the probability and consequence components of the risk.	Agree. NNPP provided similar comment. Comment no longer applicable—table format has been standardized in Rev 0C (AS).

4. PA (Palmer Vaughn)	General	Numerous new documents relating to postclosure science are being planned for the future, presumably by BSC, as reflected in the tables of "Supporting Products to be Referenced in This Section." Many of these are planned for FY07 and FY08 and TWPs for some have not been written yet. It is not clear why BSC is planning to produce these documents, as they are part of the Lead Laboratory scope of work and they are planned for well after the end of transition.	OK. If the author is requesting clarification in document text, the comment should specifically request it. Now that the Lead Lab is responsible for th 2. Page 0 of 17 of these documents should be re-considered to ensure that their preparation coincides with Lead Lab planning (AS).
5.	General	The Project Team listing format is a good way to record all participants in a modeling area, including legal, legal support, expert witnesses, etc. It would be helpful for the LL Licensing Department to have access to the database within which these forms are kept.	OK. This comment does not require any changes to the CDR (AS).
6.	General	It is recommended that the Risks portion of each modeling area (section X.6) be collected into a database that is accessible by the LL Licensing Department.	OK. This comment does not require any changes to the CDR (AS).
7.	Pg. iii	Andrew Orrell is incorrectly listed as the "SNL Manager, Lead Laboratory Licensing Department." This should be changed to "Lead Laboratory Senior Manager."	OK. This change has already been made in Rev 0C (AS).
8. AS	Pg. xiii	Definitions are missing for AWP, EBSRTA, and S&A. These should be supplied.	Agree.
9. PA (Palmer)	Sect. 2.2.5, Pg. 7 Now Sect. 3.2.5, Pg. 20 in Rev 0C (AS)	As noted in this section, this document is based on the development of the BSC Annual work Plan in response to supplemental DOE guidance, resulting in a number of changes to the technical work scope. The Lead Laboratory should be involved in the development of the postclosure portion of this work scope, since it will be transitioned to the Lead Laboratory at the end of the transition period (see also comment 4 above).	OK. This paragraph needs to be re-written to reflect Lead Lab ownership (AS).
10. PA (TSPA)	Sect. 2.2.5.1, Pg. 7, last Par. Now Sect. 3.2.5.1, Pg. 20, last Par. (AS)	An explanation should be given to justify not including these "supplemental" runs into the licensing base-case model reports.	Agree.
11. AS	Sect. 2.2.5.1, Pg. 7, last Par. Now Sect. 3.2.5.1, Pg. 20, last Par. (AS)	It is recommended that the statement that begins with "To address the perceptions by the IVRT and others…" be changed to read "To address the concerns of the IVRT and others…" There was a specific rationale behind each of the comments made by the IVRT.	Agree.

12. PA (Cliff Ho)	Sect. 2.2.5.2. Pg. 8, 2nd sent. Now Sect. 3.2.5.2. Pg. 21, 2nd sent. (AS)	This sentence is incomplete, making its meaning unclear. It should be re-written for clarification.	Agree. The corrected sentence will be: With the objective of eliminating reliance o [illegible] the maximum extent practical while not pre [illegible] the new infiltration maps will be propagated through the unsaturated zone (UZ) flow and transport models <u>but not completely through all of the downstream models.</u> 2. Page 0 of 17
13. AS	Sect. 2.2.5.3. Pg. 10, 1st Par. Now Sect. 3.2.5.3. Pg. 23, 1st Par. (AS)	This paragraph states that draft LA will be submitted to DOE HQ in the November 2007 timeframe. Is this still consistent with latest LA submittal schedule?	NNPP also commented on this. The date should be February 2008.
14. PA (Swift)	Sect. 2.2.5.3. Pg. 10, 2nd Par. Now Sect. 3.2.5.3. Pg. 23, 2nd Par. (AS)	This paragraph discusses the potential problems with using the TAD design specifications as a proxy for an actual design and notes that management of this challenge requires immediate interaction with NRC senior management. Have these interactions begun yet? The Lead Laboratory agrees that this could be a major future postclosure problem that needs to be addressed immediately.	Agree.
15. PA (TSPA)	Sect. 2.2.5.4. Pg. 11, 1st Par., last sent. Now Sect. 3.2.5.4. Pg. 24, 1st Par., last sent. (AS)	The sentence is not clear, as the syntax is off, and needs to be clarified. Also, the suggested approach seems not to take into account transparency requirements.	Agree.
16. AS	Sect. 2.2.5.4. Pg. 11, last Par. Now Sect. 3.2.5.4. Pg. 24. Last Par. (AS)	Future references to the IVRT should be avoided since the IVRT was disbanded in the spring of 2006. For example, the first sentence of this paragraph states "The most significant risk involves the potential for the IVRT to conclude that the eventually revised 2007 TSPA model is not valid for its intended use." The former IVRT would not be reviewing the next iteration of the TSPA. Same comment for other places in the document where future references are made to the IVRT.	Agree

17. PA (Palmer Vaughn)	Sect. 2.6, Pg. 15 Now Sect. 3.6, Pg. 28 (AS)	It is unclear what the difference is between BSC-R-0005 and BSC-R-0006. Also, the cryptic entries in this table (e.g., CRIT-05, SS4, WP-1, LA-2001-013 need more information/clarification so that the reader can better understand what the risks are.	Agree. 2. Page 0 of 17 First part of comment on difference betwee[illegible]-R-006 still applies; however second part of comment no longer applies because the requested clarifications have been made in this revision (Rev 0C) (AS).
18. PA (TSPA)	Sect. 6.2.5, Pg. 53 (AS) Now Sect. 7.2.5, Pg. 66 (AS)	The next to the last paragraph states, "The replacement model will be based on a conceptual model similar to that used in BSC(2004[DIRS17007])." It would be helpful and would increase the transparency of this CDR if a brief description of this conceptual model were included so that the reader would only have to refer back to the reference for more detail/additional information.	Agree.
19.	Sect. 8.6, Pg. 105, BSC-R-0010 Now Sect. 9.6, Pg. 115, BSC-R-0010 (AS)	The discussion of this risk and others in the risk tables focuses on the differences of opinion between BSC and the IVRT and Lead Laboratory with respect to the appropriate approach and methodology for a "next generation performance assessment," including the use of conservatisms, treatment of uncertainty, etc. It is recommended that this discussion be re-written to focus only on the potential risks to schedule that implementation of the Lead Laboratory approach may incur rather than amplifying differences in performance assessment philosophy between BSC and the Lead Laboratory. Focussing on these philosophical differences in documents that may become public, such as this one, is counterproductive and has the potential to negatively impact the licensing process. Same comment for other places where similar discussions are presented, including the discussions on pages 107 and 168.	OK. This comment is no longer applicable because this language has been removed from the risk tables and elsewhere in the document in Rev 0C. (AS)
20.	Sect. 14.6, Pg. 268, BSC-R-0002 Now Sect. 15.6, Pg. 266, BSC-R-0002 (AS)	The description of this risk states that SAR Section 2.3.9 (Saturated Zone Flow and Transport) provides the basis for "screening of FEPs related [to] UZ Flow and Transport." Shouldn't this be SZ Flow and Transport?	Agree. Latest schedule shows SAR Section 2.3.9 addresses the Saturated Zone. This comment is no longer applicable because this language has been removed from the risk table. (AS)
21. AS	General	References to "peak dose" should be replaced with something more descriptive. If "peak dose" occurs after 1,000,000 years is it proposed to be calculated to whatever time is required?	Reference to "peak dose" should be replaced with reference to "peak dose within 1,000,000 years."

22. PA (TSPA)	General	The list of supporting products describes changes being made to certain documents and identifies other documents where no changes are being made. For the latter there is no indication as to why no changes to certain documents are required. It is not always obvious why this is the case so if this table is to be retained it is suggested that it be made more comprehensive by including reasons why no changes are required to those documents.	Agree. 2. Page 0 of 17
23. PA (Palmer Vaughn)	Section 2.2.1, pg. 4, 3rd par., last sentence Now Sect. 3.2.1, Pg. 17, 3rd Par., last sentence (AS)	It is not clear what the significance or purpose is of the sentence "To the extent practical, no NRC-sponsored research or analyses are discussed or presented, unless that information is used to support the technical basis or is different from the technical basis, and the reasons for the differences need to be present." Please clarify.	Agree. Recommend sentence be deleted.
24. PA (TSPA)	General	The phrases "best-estimate" and "realistic" should be replaced with the tems "performance margin analysis (PMA)" and "next generation performance assessment", as appropriate. PMA is intended to be a performance assessment using the TSPA architecture but with less bias in the characterization of uncertainty. It will be used to evaluate the performance margin associated with the compliance baseline in the summer 2007 time frame. The next generation performance assessment is not confined to the current TSPA architecture and is intended to incorporate state of the art algorithms, computational hardware, and new information as available. The next generation PA will support license defense activities after the 2008 LA submittal.	No comment.
25. PA (TSPA)	Section 2.2.5.1, pg. 8, 2nd par. Now Sect. 3.2.5.1, Pg. 21, 2nd Par. (AS)	BSC and SNL should discuss Section 2.2.5.1 to make sure it is consistent with the approach that will be taken under the Lead Laboratory. For example, it is not clear what is meant by the following " The overall philosophy embodied in the scope of work for each technical work area is intended to ensure the postclosure performance assessment adequately incorporates the key aspects impacted by the changes and, as a first priority, to address any potential optimism in the TSPA. This approach uses sensitivity and impact analyses and supplemental calculations directly in the licensing basis without propagation through the TSPA." This does not seem to coincide with the Lead Laboratory approach.	Agree.

26. AS	Section 2.2.5.1, pg. 8, 4th par. Now Sect. 3.2.5.1, Pg. 21, 4th Par. (AS)	This paragraph seems to imply that part of risk control would be to limit the use of new information after submittal, as suggested by the sentence "Part of the strategy to address these challenges would be to determine the appropriate time to introduce any new information". It is suggested that this approach be reconsidered. It would seem to be more appropriate, as well as more acceptable to the NRC and stakeholders, for new information to be evaluated and utilized as it becomes available rather than when it is convenient or advantageous to do so.	OK. 2. Page 0 of 17
27. PA (Cliff Ho)	Section 2.2.5.2, pg. 8, par. 1, sentence 2 Now Sect. 3.2.5.2, Pg. 21, Par. 1, Sentence 2 (AS)	This section should be discussed between BSC and SNL to ensure that the Lead Laboratory approach is reflected here. For example, the underlined portion needs specific discussion.	OK. The current plan is to propagate the new infiltration results through the UZ flow, Calibrated Properties, UZ Transport, and UZ transport abstraction models to TSPA. Other impacts such as drift seepage and SZ will be handled via sensitivity analyses. Current SNL work scope includes tasks under AUZM22 (WS8) to assess the impact of infiltration changes on Drift Seepage, SZ recharge, and SZ flow and transport abstraction. These sensitivity studies will be documented in a separate report. The last 2 bullets in Section 3.2.5.2 are no longer valid due to DOE acceptance review directed changes to TWP-MGR-PA-000036 Rev. 02. Instead of reducing reliance on calibrated properties, we are directed for consistency to use updated calibrated properties prepared as part of revision to the UZ flow model. Accordingly, please change these 2 bullets as follows: • Revise the Multiscale AMR to use recalibrated one-dimensional drift-scale UZ hydrologic properties, and to treat percolation flux parametrically rather than as a direct feed from the UZ flow and infiltration models. The latter objective will be accomplished by performing thermal-hydrologic calculations for a range of percolation flux conditions and interpolating the results to address percolation values used in TSPA. (SAR Section 2.3.5) • Revise the THC Seepage Model AMR to use updated UZ hydrologic properties, consistent with the Multiscale model, and to use new flux values consistent with updated estimates of infiltration and percolation. In addition, the THC model will be revalidated against the Drift-Scale Test, with appropriate numerical sensitivity analyses added (SAR Section 2.3.5).

28. PA (Palmer)	Section 2.2.5.3, pg. 10, bullets 2 and 4 Now Sect. 3.2.5.3, Pg. 23, bullets 2 and 4 (AS)	These bullets or parts of these bullets do not relate to TADs and should be moved to the appropriate section.	OK. 2. Page 0 of 17
29. PA (TSPA)	Section 2.2.5.3, pg. 10, bullet 5 Now Sect. 3.2.5.3, Pg. 23, bullet 5 (AS)	This bullet does not appear to belong in this list. Also, BSC and SNL should discuss the content of this bullet to ensure that it is consistent with Lead Laboratory planning.	The next-to-last bullet in Section 3.2.5.3 should be changed as follows: • Complete Postclosure Thermal Envelope Study to show that the postclosure temperature limits can be accommodateed for a wide range of repository thermal loads including TAD canisters, represented by a selected Estimated Limiting Waste Stream. This work will provide the assessment of hydrogeologic, geochemical, and geomechanical effects to the range of anticipated thermal loadings, that is required by 10CFR63.21 (SAR Section 2.3.5).

30. PA (TSPA)	Section 2.2.5.4. pg. 12, par. 1, bullets 1, 5, and 6. Now Sect. 3.2.5.4, Pg. 25, Para. 1, bullets 1, 5, and 6 (AS)	BSC and SNL should discuss these to ensure they are consistent with Lead Laboratory planning.	Some of the following information pertains to the EBS RTA which is now a responsibility of TSPA, so Bob and Dave will need to weigh in. 2. Page 0 of 17 Firstly, bullet 1 needs to be deleted if we are not going to MINC in the UZ flow and transport models. Even if we are, it is not clear that the EBS-UZ interface model needs to updated. Bullet 4 is not quite right—I recommend the following replacement: • Revise Drift-Scale THC and in-drift chemistry models to address unresolved IVRT comments related to selection of water composition data for modeling, variability in seepage water chemistries, chemical binning of potential seepage waters, and parameter uncertainty in THC models. (SAR Section 2.3.5) Similarly, bullet 5 should be revised: • Revise EBS Physical and Chemical Environment model to reevaluate the statistical uncertainties used for in-drift chemistry to improve realism, and to simplify the in-drift chemistry abstractions used in TSPA to represent seepage and invert porewaters (SAR Section 2.3.5). Bullet 6 goes a bit too far and should say: • Demonstrate integration of the in-package models (waste form, chemistry, and radionuclide transport) by exercising them in a TSPA environment and evaluating reactant consumption, in-package humidity, and other effects as potential conservatisms (SAR Section 2.3.7).
31. PA (TSPA)	Section 2.2.5.4, pg. 13, par. 1, bullet 1. Now Sect. 3.2.5.4, Pg. 26, Para. 1, bullet 1 (AS)	BSC and SNL should discuss to ensure this text, especially the reference to 1 cell versus 2 cell, is consistent with Lead Laboratory planning.	See previous response.
32. PA (TSPA)	Section 4.2.5 and 4.2.6 Now Sect. 5.2.5 and 5.2.6 (AS)	What is meant by "XX?"	Agree. It is likely "XX" means "Reserved", which should be used instead. Need to determine what should go into these sections and put it in. (AS)

33. PA (Cliff Ho)	Section 6.2.5 Now Sect. 7.2.5 (AS)	BSC and SNL should discuss this section, as it does not appear to be consistent with Lead Laboratory planning.	OK. The paragraphs in Section 7.2.5 beginning with "Two independent parallel paths are…" and continuing through the next three paragraphs and ending with " … relat 2. Page 0 of 17 tch approach." Should be deleted. Rather, this section should indicate that the infiltration model is being revised and the changes to infiltration estimates from the revision will be propagated through the UZ models. Two additional models are being revised in part because of changes to infiltration. These additional two are: Radionuclide Transport Model under Ambient Conditions (MDL-NBS-HS-000008) and Particle Tracking Models and Abstraction of Transport Processes (MDL-NBS-HS-000020). The bullet indicating that the Calibrated Properties Model will be revised is incorrect and should be deleted.
34. PA (Cliff Ho)	Section 6.2.5, page 54, par. 1, last sentence. Now Sect. 7.2.5, Pg. 67, Para. 1, last sentence (AS)	What is intended by the following sentence? Please be more specific. "The criterion for management selection of the approach to use for simulating infiltration is the relative technical merit of each approach." The reviewer was under the impression that the Lead Laboratory has already selected the Massif approach.	Agree. . The paragraphs in Section 7.2.5 beginning with "Two independent parallel paths are…" and continuing through the next three paragraphs and ending with " … relative technical merit of each approach." Should be deleted. Rather, this section should indicate that the infiltration model is being revised and the changes to infiltration estimates from the revision will be propagated through the UZ models. Two additional models are being revised in part because of changes to infiltration. These additional two are: Radionuclide Transport Model under Ambient Conditions (MDL-NBS-HS-000008) and Particle Tracking Models and Abstraction of Transport Processes (MDL-NBS-HS-000020).
35. AS	Section 7.6 , Section 13.6 , Section 14.6 Risks, item BSC-R-0002 Now Sects. 8.6, 14.6, and 15.6, BSC-R-0002 (AS)	Consider re-wording the following sentence: "TSPA modifications due to Independent Validation Review Team Delay Submittal of the License Application (TSPA-06) – Because to this section provides the basis for models supporting TSPA and the screening of FEPS related UZ Flow and Transport" The meaning is unclear and needs clarification. Additionally, the delay is not a result of the IVRT review but a result of needed improvements.	Agree. Partially resolved by BSC's elimination of 2nd portion in Rev 0C (AS).

36. PA (TSPA)	Section 7.6 . Section 13.6, Section 14.6 Risks, item BSC-R-0028 Now Sects. 8.6, 14.5, and 15.6, BSC-R-0028 (AS)	This item seems out of place in this list. It appears to be based on a pre-conceived notion of what is meant by "Alternative TSPA Approach". Please define what this approach is. BSC and SNL should discuss this to ensure that it accurately reflects Lead Laboratory planning.	Agree. May need to refer to original version (Rev 0 [illegible] BSC approach, since no elaboration is given in c [illegible] 2. Page 0 of 17
37. PA (Cliff Ho)	Section 8.2.5, pg. 90, par. 1, s 2. Now Sect. 9.2.5, Pg. 102, Para. 1, sent. 2 (AS)	It is not clear why the ambient seepage models and the thermal-hydrologic seepage models are not scheduled for revision when infiltration is expected to change. Isn't seepage influenced by infiltration? Please provide an explanation for this apparent discrepancy.	Agree. The planned recalibration of the UZ Flow model and submodels will address the impacts of changes in infiltration. Preliminary results indicate that infiltration for the calibrated cases will be greatly different from the current case. One would then expect that seepage for those cases will differ greatly from the present case. Therefore it is reasonable to assess the impact to seepage and TH as sentivity studies as are currently planned. This is a risk, but one that is believed to be small and manageable.
38. PA (TSPA)	Section 8.6, Section 10.6 Item BSC-R-0026 Now Sects. 9.6 and 11.6, BSC-R-0026 (AS)	BSC and SNL should discuss this item to ensure it is consistent with the Lead Laboratory approach.	OK. May need to refer to original version (Rev 0A) to determine BSC approach, since no elaboration is given in current version. (AS)
39. PA (Palmer)	Section 8.6, Section 10. Item BSC-R-0028 Now Sects. 9.6 and 10.6, BSC-R-0028 (AS)	LL needs to review and develop this text. In particular, LL takes exception to the statement that "This risk is potentially exacerbated by transition of the postclosure science activities from BSC to Sandia." Etc. Statements about an "Alternative TSPA Approach" appear to be somewhat generic and speculative. Suggest this entire section be eliminated or some more informative statements concerning alternative TSPA approaches be developed through discussions between SNL and BSC.	Agree. Recommend deleting the sentence. Agree. May need to refer to original version (Rev 0A) to determine BSC approach, since no elaboration is given in current version. (AS)

40. PA (TSPA)	Section 10.3, pg. 157, Item DIRS 175058 Now Sect. 11.3, Pg. 165, Item DIRS 175058 (AS)	The date of June 2007 is given as the date for submittal of new information supporting a new screening decision for dust deliquescence. Since feeds to TSPA are needed by 3/20/07, this date will preclude its use in TSPA. This should be made clear.	Agree. 2. Page 0 of 17
41. PA (Palmer)	Section 11.2.5, Page 178, par. 1, last sentence. Now Sect. 12.2.5, Pg. 180, Para. 1, last sent. (AS)	"However, the postclosure performance issues may be limited and may have only a minor impact on this section of the SAR." This sentence is then followed by 12 pages of changes, supporting documents, and risks. This appears to imply that more than a minor impact is already identified. Suggest the sentence be removed.	Agree.
42. PA (TSPA)	Section 11.2.6, pg. 181 Now Sect. 12.2.6, Pg. 183 (AS)	There are a number of other TADs related impacts that are not identified, such as changes to chemistry, sorption capacity onto corrosion products, criticality, changes to the footprint, waste loading, inventory, etc.	Agree.
43. PA (TSPA)	Section 12.2.5, pg. 202, par. 5 Now Sect. 13.2.5, Pg. 203, Para. 5 (AS)	BSC and SNL need to discuss to ensure that this paragraph reflects the current Lead Laboratory approach.	OK. The approach set forth in TWP-MGR-PA-000020 Rev. 03 is much less extensive than that represented in the draft CDR. The EBS RTA is now a responsibility of TSPA so Bob or Dave will want to fill in details. I recommend the paragraph be revised as follows: Sections 2.3.7.6 through 2.3.7.8 regarding fuel and cladding degradation, are expected to change. The abstractions for degradation of in-package materials will be revised to evaluate cumulative consumption of reactants used in degradation of in-package materials. Amounts of reactants such as H_2O and O_2 can then be compared with separately estimated limitations on mass-transfer into breached waste packages, to evaluate possible conservatisms in the degradation models.

44. PA (TSPA)	Section 17.2.5 Now Sect. 18.2.5 (AS)	More information should be provided outlining the changes anticipated for TSPA and the need for those changes.	Agree. 2. Page 0 of 17
45. PA (Tom Pfeifle)	Section 17.6 Now Sect. 18.6 (AS)	No risks are identified. Suggest providing a list of risks as in other sections.	Agree.

CONCEPT OF OPERATIONS for the YUCCA MOUNTAIN PROJECT TECHNICAL DATA MANAGEMENT SYSTEM

Version: 1.0

March 31, 2007

Prepared by:

Sunita Moonka, 04519, 284-3259, smoonka@sandia.gov

Jason Follingstad, 04538, 844-66623, jtfoll@sandia.gov

Sean Hendrickson, 04515, 284-7887, slhendr@sandia.gov

Brian Scott, 04531, 844-4762, bcscott@sandia.gov

Walter Walkow, 04538, 844-3750, wmwalko@sandia.gov

This page was intentionally left blank.

Signature Page

Raymond W.. Shaum

Signature Date

Stanley A. Orrell

Signature Date

Revision History

Revision Date	Revised By	Comment
11/13/06	Sunita Moonka	Initial template
12/14/06	Patricia Hackney	Format, edits, insert new inputs from Sean and Sunita
12/15/06	All team	Review Section 3.
12/18/06	All team	Review Section 3. New and revised graphics (business process, fact models, data flows)
12/19/06	Sunita Moonka	Revised graphics business process and user/agencies; additional terms
12/19/06	Sean Hendrickson	Revised subsection (out of scope systems); first draft current state issues; first draft "to/be" state
12/19/06	Patricia Hackney	Insert new/revised material into version 0.1[5]
12/20/06	Patricia Hackney	Format revisions (tables xx and xx); edits (version 0.1[5b])
01/05/07	Sunita Moonka	Added DIRS process, YMP/SNL Logo
01/15/07	Patricia Hackney	Revisions from the LV 1st review
01/23/07	Patricia Hackney	Ray Shaum: general revisions
01/27/07	Patricia Hackney	Ray Shaum: cite QA references; move issues to appendix
02/07/07	Patricia Hackney	Replace Figure 2; insert new figure in Section 7.0; draft Section 6 (issues and risks)
02/14/07	Patricia Hackney	Revise section about TSPA, revise, alphabetize terms; edit issues appendix
02/16/7	Patricia Hackney	Include changes from Sunita; new and revised graphics; create appendix for graphics; draft acknowledgements
02/17/07	Sunita Moonka	Inserted new material
02/19-20/07	Patricia Hackney	Revise Section 1/4 per discussions with Sunita.; move all figures to appendix
02/23/07	Sunita Moonka	Inserted new material and Migration Issues provided by Walter Walkow
O3/06/07	Sunita Moonka	Added new material
03/07/07	Sunita Moonka	Added make buy analysis done by Brian Scott and Jason Follingstad
03/14/07	Sunita Moonka	Updated based on TDMS Team feedback
03/16/07	Donna Alexander	Format, edits, insert new inputs from Sunita
03/16/07	Walter Walkow	Updated acknowledgments
03/17/07	Sunita Moonka	Updated Julies Warner's comments
03/17/07	Sunita Moonka	Updated Rob Rechard's comments

03/17-27/07	Donna Alexander	Format, edits, Insert new inputs from Sunita

Acknowledgements

The authors would like to thank the following individuals for their valuable advice, comments, reviews, and contributions in preparing this document. However, in some cases, key personnel were unable to provide requirements because of tight schedules. Every attempt to get their input was made while adhering to the agreed upon schedule.

Name	Involvement Rational	Expertise	Title/Role/Class
Ray Shaum	Management overview	Management	Manager (SNL 6786)
Duane Allred	Conduit	Agencies external to LL	SNL 6778
Ed Miller	QA Review	CPR process	SNL 6786
Rob Rechard	Review	Performance assessment	SNL 6783
Mike Russell	CR advice	CR	SNL 6786
David Seamans	Query, reports, ATDT capability requirements, etc.	ATDT data entry, knowledge of inputs, outputs for ATDT	Tech DB processor (SNL 6786)
C. J. Houston	Query, reports, ATDT capability requirements, external interface requirements, etc.	GI data entry, external organization coordinator	Tech DBA (SNL 6786)
Rick Maul	Current database technology infrastructure	BSC IT database administration	BSC IT DBA Manager
Connie Kawasaki	Sole source for DIRS, SME	DIRS	DIRS Admin (SNL 6786)
Matt Hankins	Current technology infrastructure; validation of current technology infrastructure	BSC IT Supervisor of DIRS and ATDT	BSC Scientific Systems Supervisor (SNL 17461)
Mary Alena Martell	Overall systems and business perspective, validation of phase deliverable	Business perspective	TDM Team Lead and Data Qualification Team Lead SNL 6783)
Mike Jaeger	Overall systems and business perspective, license defense perspective, validation of phase deliverable	Business perspective	(SNL 67783)
Aaron Engel	BSC IT applications developer	DIRS, ATDT applications knowledge	BSC IT applications developer
Salvatore Marano	BSC IT applications developer	DIRS, ATDT application knowledge	BSC IT applications developer
Kim Denton-Hill	Management	Information systems engineering	Manager (SNL 4519)
Dawn Abbott	Engineering management	Engineering management	(SNL 4511)
Alan Armentrout	Peer review	Information model	
Wendy Shaneyfelt	Cognitive sciences and peer review	requirements and use cases	(SNL 6341)
David Clifford	Management	Information engineering	Manager (SNL 4514)
Van Pham	Data warehouse	Data warehouse	Enterprise data warehouse project lead (SNL 4512)
Brad Mancke, Sue Medeiros and Damon Gerhardt	Data and text mining	DTM recommendations for TDMS	SNL Org 4511
Patricia Hackney	Technical writer	Technical writing, editing	Sr. Technical writer

			(SNL 4512)
Donna Alexander	Technical writer	Technical writing, editing	Technical writer (SNL 4512)

Table of Contents

List of Figures and Tables

List of Tables

Executive Summary

In 2002, the President recommended and the Congress concurred with that recommendation that the U.S. Department of Energy (DOE) construct a deep underground repository for the permanent disposal of spent nuclear fuel and high-level radioactive waste at Yucca Mountain in Nevada. Currently, the Yucca Mountain Project (YMP) of the DOE is preparing a license application to the U.S. Nuclear Regulatory Commission (NRC) to construct the repository. DOE plans to submit the license application (LA) in July of 2008. NRC regulations require reasonable assurance that the completed repository meet Environmental Protection Agency (EPA) environmental radiation protection performance standards.[3] To this end, YMP conducts a performance assessment (PA) to simulate possible behavior of the repository. Scientific and engineering data and information used in the PA must have pedigree; hence, the management of this data and information is critical to the credibility and accountability of the simulations. In 2006, DOE assigned Sandia National Laboratories (Sandia) as the Lead Lab to provide scientific expertise to conduct the PA and support defense of the license application.

The Lead Lab has undertaken several evaluations of the current state of affairs of the infrastructure that supports the PA. This document presents the analysis performed by the Information Systems Center (4500) of Sandia, in its roll as YMP Lead Laboratories of the existing technical data management (TDM) process that supports the YMP PA. This analysis complied with the governing DOE regulations and orders[1] as well as Sandia Corporate requirements and guidelines[2]. (See [redacted] for details about the government and Sandia Corporate regulations, orders, requirements, and guidelines.)

The TDM process manages essential scientific and engineering data regarding the site characterization and licensing application for the Yucca Mountain repository. TDM processes were implemented for both ensuring the integrity and quality of this YMP data and maintaining traceability for references in legally required, government-deliverable YMP documents that are to be accessible to the public.

Currently, two systems comprise the TDM process, the Technical Data Management System (TDMS) and the Document Input Reference System (DIRS), hereafter referred to collectively as the TDM Systems. The analysis of the TDM Systems included the requirements for a proposed replacement system for the existing TDMS and DIRS to support the license defense stage of the license application for the repository. The relationship of the TDM Systems to other YMP analysis efforts, processes, and procedures was considered to ensure that all impact aspects of the current systems and potential replacement systems were understood and analyzed.

The existing TDM Systems are a collection of six major stand alone databases, user interface screens, and processes requiring extensive manual manipulation. The TDM is complex because of the numerous applications used by TDMS and DIRS. (See [redacted] for details of the current information systems architecture.)

TDMS interfaces with at least seven other applications to support the input of technical data by the authors, the creation and input of metadata and indexing information associated with the

technical data, and search and access to the information by authorized users. The TDMS is organized around three main functionalities (i.e., data entry, data quality assurance, and data retrieval/usage). These involve five separate sets of roles and responsibilities (i.e., originator, reviewer, data coordinator, database administrator, and records coordinator) as depicted in [redacted]. (See Appendix A.)

DIRS, in conjunction with multiple applications, is primarily a reference management system that supports the formatting of bibliographies and cited works, and cross-references document inputs and products to allow tracking of these references. DIRS is organized around three main functionalities (i.e., reference entry, reference verification, and reference usage). These involve three separate sets of roles and responsibilities (i.e., originator, reference locator, and DIRS administrator) as depicted in Figure 6(see Appendix A). (See Section 4.0 for details of the current TDM Systems.)

We found, serious issues and gaps in the TDM (see [redacted]) in our analysis. The TDM Systems do not automatically suppor,t and in some cases inhibit,t the flow of the work. By not automatically supporting the flow of work, humans must manually ensure the integrity, accountability, and traceability of the data. These issues and gaps include:

- Suboptimal business processes (e.g., no IRAN process for QSL data in TIC, no time limit on IRAN response, less than optimal quality control on USGS data submitted directly into RPC)
- Parts of the business processes are supported by TDMS, DIRS, and other peripheral systems while critical processes (e.g., impact review assessment notification, submission of technical products and product references, quality control, review of technical data, tracing developed data to source data) are accomplished manually.
- Most TDMS operating system software, middleware, database management system software, and programming languages are dated and are often unsupported technologies on the Bechtel SAIC Company (BSC) network.
- Extensive manual manipulations are necessary to accomplish many of the operational procedures, which is time consuming and labor intensive, especially if errors are to be avoided.
- Each of the functional areas has supporting applications operating in a legacy infrastructure environment consisting of "stovepipe" systems and data.
- There are security and maintenance issues. For example, by design of the system, it is necessary for TDMS administrators to have full access to the file server and production database so that they can publish the static web pages, upload datasets, and update the database when they receive new or changed datasets. Because of this, administrators have the ability to accidentally manipulate production data without going through the application, thus bypassing access controls.

Formatted: Bulleted + Level: 1 + Aligned at: 0" + Tab after: 0.25" + Indent at: 0.25", Tabs: 0.06", List tab + Not at 0.25" + 0.63"

Recommendation for moving forward

We recommend that the current TDM System be replaced. The replacement system must automatically track data items through the system from end-to-end; conclusions developed and

published for the Licensing System must be able to automatically verify how data was developed throughout the analysis and modeling process; and referential integrity must be maintained by the database system to ensure the consistency and accuracy of the data.

The goal is to create a streamlined optimal exchange and common understanding among various organizations and agencies that implement specific areas and to rid the process of duplicated efforts and manual manipulations. Enterprise Business Modeling and Value Stream Analysis is recommended to identify business areas that are either not addressed or are weak. This approach will also help the IT Integration team target and prioritize business areas that need automation. Individual projects can then be evaluated with an understanding of how their effort fits into the overall business.

Redevelopment of DIRS and TDMS would provide the following desired changes. (See [redacted] for a complete analysis of desired changes and recommendations.)

- Overhauled longstanding outdated technology
- Reduced manual procedures (e.g., checking the accuracy and validity of data and references, change history, access control, and trace development)
- Integrated corresponding systems supporting SIP (e.g., TDMS, CDIS, RIS, TIC, SCM and CSITS)
- Enhanced data quality and integrity
- Enhanced system security and maintainability (e.g., access control and backups)
- Enhanced reporting capability.
- Integrated Lotus Notes functionality into TDMS and DIRS
- Streamlined TDMS process with minimum redundancies (see Figure. 1 in Appendix A for the to–be TDMS process)

Existing Sandia programs, hardware, software, and systems, (e.g., DART SDM, AMR Exploration Tool, Issue Management Database and SDDB) as well as COTS products, (e.g., WebPE, SharePoint, Documentum, Stellent, Oracle, and Serena Team Track) were considered as potential solutions to the issues and risks that were identified

A Make/Buy and Reuse analysis (see Section 8.0 for a detailed analysis) shows that the best strategy would be a ***combination of in-house development, while leveraging the re-use of Sandia tools (e.g., DART and AMR Exploration Tool),and COTS*** (Serena Team Track, e-matrix, Oracle database and APC server). The advantages include:

- Delivering custom applications built to fit the YMP Business process and goals rather than modifying YMP business processes around the packaged software solution
- Existing tools can be re-used lowering overall development cost while improving quality
- Instead of one big drop after development, the possibility of smaller, iterative drops for user testing and feedback with the in-house staff providing tighter control of the project
- Levering multi-application data integration can require significant ongoing administrative effort, which can be reduced with custom tools or reports

- Control over delivery schedules especially given the short deadline (the next license application is scheduled for 2008).

The consolidation of TDMS and DIRS was considered but then dismissed because mapping of the functional areas supported by DIRS and TDMS does not show an overlap between the two systems. They support unique business functions: DIRS supports Document References, whereas TDMS supports Data Management. Consolidation of the two systems would only make sense if all other peripheral systems (e.g., RIS, TIC, CDIS,) that support SIP were also included in that consolidation. To accomplish that, further analysis of the peripheral system and re-engineering of business processes with the objective of achieving more efficiency has to occur first. Also, because a single consolidation implementation must support the needs of many agencies (OCRWM wide), it must incorporate agility, flexibility, compatibility, and ease of use to ensure that it is adapted quickly, seamlessly, and with few, if any, service disruptions. We do not see this as a possibility by implementing one monolithic system in the current environment.

The above TDM Systems changes and recommended solutions address the issues and risks identified. However, the redevelopment effort has number of Challenges and Constraints including:

- Data migration issues are significant. Data migration challenges including finding inconsistent, duplicated, and orphaned data, (For a complete overview of data migration issues and recommendations, see [redacted].) Sound practices must be followed:
  - Analyze data to identify data that can not be converted (bad data values, etc.)
  - Obtain corrections from data owners where possible
  - Employ data cleansing procedures to correct data
  - Develop migration scripts that can be repeated
  - Employ migration tools to assist in conversion (e.g., SQL Server SSIS or DTS, Data Stage – data warehouse tool)
  - Develop auditing processes to account for all data converted/not converted
  - Document the migration process
  - Test the migration process
  - Review audit results and migration process with data owner
- Applications deployed/managed by Sandia IT must follow Sandia IT standards for application development methodology, development tools, database management, and operations (For complete IT Infrastructure standards see [redacted].)
- The current Sandia YMP environment offers considerable challenges since it does not provide the required infrastructure; its support now is for basic services (e-mail, SharePoint, and desktop office products). There is no plan to provide a complete infrastructure for development and deployment, ([redacted] for a list of required infrastructure.)
- Since the deployment of new DIRS and TDMS on the Sandia network will still require access to data and applications remaining on Bechtel network, managed by Bechtel IT, access to the Bechtel database is critical and could be a challenge. Receiving access to the data and staff must be negotiated upfront before redevelopment work begins.

Formatted: Bulleted + Level: 1 + Aligned at: 0" + Tab after: 0.25" + Indent at: 0.25", Tabs: 0.06", List tab + Not at 0.5"

Formatted: Indent: Left: 0", Hanging: 0.31", Bulleted + Level: 1 + Aligned at: 0" + Tab after: 0.25" + Indent at: 0.25"

- Customer involvement is always a challenge. Since the next phase of the TDMS/DIRS development will require significant involvement with the TDMS/DIRS user and administration community for testing and user interface design validation it will be critical to negotiate their involvement upfront.

Formatted: Indent: Hanging: 0.19", Bulleted + Level: 1 + Aligned at: 0" + Tab after: 0.25" + Indent at: 0.25"

1.0 INTRODUCTION

This document addresses an assignment from the Department of Energy (DOE) to Sandia National Laboratories (Sandia), in its roll as the Yucca Mountain Project (YMP) Lead Laboratory. The task is to complete a requirements analysis of a replacement system for the existing technical data management (TDM) process that supports the YMP scientific investigation process. This document is the first in a network of documents (e.g., Project Proposal, Project Plan, and Software Requirements Specification Document) for the subject analysis.

Two systems, the Technical Data Management System (TDMS) and the Document Input Reference System (DIRS), comprise the YMP TDM process and are referred to collectively as the TDM Systems. The TDM Systems analysis efforts complied with the governing DOE regulations and orders[1] as well as Sandia Corporate Process Requirements (CPRs) and guidelines.[2] Additionally, the Serena® Requirements Traceability Management (RTM) repository was used for capturing and managing requirements identified during the analysis. (See [redacted] for an in-depth discussion of these regulations, orders, CPRs, and guidelines.)

The YMP uses the TDM Systems to manage essential scientific and engineering data regarding the site characterization and licensing application for the Yucca Mountain repository. The TDM Systems objectives are to ensure the integrity and quality of this YMP data and to maintain traceability for references in legally required, government-deliverable YMP documents that are to be accessible to the public.

1.1 Background

In 2002, the President recommended and the Congress concurred with his recommendation that the U.S. Department of Energy (DOE) construct a deep underground repository for the permanent disposal of spent nuclear fuel and high-level radioactive waste at Yucca Mountain in Nevada. Currently, the Yucca Mountain Project (YMP) of the DOE is preparing a license application to the U.S. Nuclear Regulatory Commission (NRC) to construct the repository. DOE plans to submit the license application (LA) in July of 2008. NRC regulations require reasonable assurance that the completed repository meet Environmental Protection Agency (EPA) environmental radiation protection performance standards.[3] To this end, YMP conducts a performance assessment (PA) to simulate possible behavior of the repository. Scientific and engineering data and information used in the PA must have pedigree; hence, the management of this data and information is critical to the credibility and accountability of the simulations. In 2006, DOE assigned Sandia National Laboratories (Sandia) as the Lead Lab to provide scientific expertise to conduct the PA and support defense of the license application.

The Lead Lab has undertaken several evaluations of the current state of affairs of the infrastructure that supports the PA. This document presents the analysis performed by the

Information Systems Center (4500) of Sandia, in its roll as YMP Lead Laboratories of the existing technical data management (TDM) process that supports the YMP PA. This analysis complied with the governing DOE regulations and orders[1] as well as Sandia Corporate requirements and guidelines[2]. (See [redacted] for details about the government and Sandia Corporate regulations, orders, requirements, and guidelines.)

1.2 Problem Statement

Currently, the TDM Systems is a collection of six major databases, user interface screens, and processes requiring extensive manual manipulation. Although current functions can guarantee that current processes are being followed, the TDM Systems cannot guarantee the "correctness" of the process nor the "correctness" or authenticity of the data, and consequently, accountability for license defensibility may fail in certain cases. Additionally, most of the TDM Systems hardware, operating system software, middleware, database management system software, and programming languages are outdated technologies. Furthermore, the requirements analysis of a replacement system must comply with both government and Sandia quality assurance requirements.

2.0 QUALITY ASSURANCE DRIVERS

2.1 DOE Office of Civilian Radioactive Waste Management

The TDM Systems must comply with the DOE Office of Civilian Radioactive Waste Management (OCRWM), Office of Science and Technology and International (OSTI) Program guidelines, standards, and requirements for research, development, test, and analysis materials and methods for use in enhancing applications. The governing documents are:

1. DOE/RW-033P, "Quality Assurance Requirements Description" (QARD), and
2. Attachment 1 "Quality Assurance Requirements for Work Authorized by OCRWM Program and Funding Guidance Memorandum."

Sandia implemented the Sandia OSTI Quality Assurance Program to address OSTI requirements. The Sandia OSTI Quality Assurance Program is implemented via the Sandia Quality Assurance Program Plan (QAPP) to satisfy the requirements of the QARD for YMP. Of particular impact on the tasks discussed in this ConOps are Sandia guidelines for establishing processes, procedures, and responsibilities in the Sandia QAAP, *Supplement V, Control of Electronic Management of Data.* The following guidelines apply to this supplement:

- IM-PRO-002, Control of Electronic Management Information
- IM-PRO-003, Software Management
- IM-PRO-005, Software Independent Verification and Validation
- IM-PRO-006, Independent Verification and Validation
- SCI-PRO-002, Records Management
- SCI-PRO-004, Managing Technical Product Inputs
- TST-PRO-003, Scientific Notebooks

2.2 Sandia Corporate Process Requirements

Additionally, updates and replacements to the TDM Systems outdated processes and technologies must comply with corporate quality assurance drivers such as the Corporate Policy

Statement Requirement 001.3. *CPSR001.3: Integrated Laboratory Management System (ILMS)*, and in Corporate Process Requirements (CPRs), such as *CPR001.3.2 Quality Assurance* and *CPR001.3.6 Corporate Engineering Excellence*. These Corporate requirements are derived from additional government guidelines and standards, such as the *DOE Order 414.1C: Quality Assurance*, and *DOE/NNSA Weapons Quality Policy (OC-1)*. Tasks identified in this document follow the Software Lifecycle (SILC) processes and procedures as the implementation of these quality assurance drivers. An example of an SILC process/tool used for this analysis is the RTM repository used for storage and management of all requirements for the project.

2.3 DOE Total System Performance Assessment

The government requirement for reasonable assurance that repository performance objectives meet EPA environmental radiation protection standards must be demonstrated in the form of a DOE performance assessment that identifies repository features, events, and processes (FEPs). A primary element of this performance assessment is the Total System Performance Assessment (TSPA). TSPA uses a scenario-development process that is based on the methodology developed by Cranwell[4] for the NRC.

3.0 3.0YUCCA MOUNTAIN TOTAL SYSTEM PERFORMANCE ASSESSMENT

Figure 1 graphically depicts the DOE mandated TSPA scenario-development process for the Yucca Mountain repository. This process is critical to understanding the role that the TDM Systems play in providing pedigreed data. This understanding, in turn, is the foundation for determining how the TDM Systems can meet repository objectives.

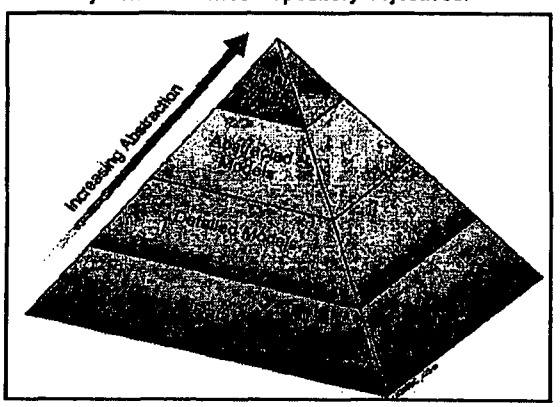


Figure 1. Steps in Developing a Total System Model

The four steps in the TSPA scenario-development process depicted in Figure 1 and how they are implemented for YMP are described below.

- **Data Level:** ***Identify FEPs potentially relevant to the performance of the Yucca Mountain repository.*** The broad foundation of the pyramid represents twenty years of accumulated knowledge, collected in the field and in the laboratory, regarding the Yucca

Mountain repository system. These data are used to identify the set of possible FEPs that may be present in the repository system.

- **Detailed Models:** ***Screen each FEP, and reach a decision to either "include" or "exclude" the FEP from further TSPA consideration.*** This stage of the TSPA includes the development and testing of models that conceptually describe the retained FEPs and their outcomes regarding repository performance. The conceptual models consist of sets of hypotheses, assumptions, simplifications, and idealizations that together describe the essential aspects of a system or subsystem of the repository relative to performance.

  An example of such a model, or interconnected models, is the description of the movement of water molecules and dissolved radionuclides by diffusive flow in rock pores or by advective flow in fracture openings in the bedrock surrounding the repository and through the saturated zone (SZ) below the repository.

  Because the TSPA process deals with future outcomes and includes uncertainty in both process descriptions and parameter values, there may be several alternative conceptual models (ACMs) that provide reasonable descriptions of a particular system or subsystem. Therefore, an essential element of the TSPA process is to capture uncertainty in probabilistic analyses that represent likely outcomes based on the best available parameter values and the processes involved. Such analyses are documented in Analysis Model Reports (AMRs) that are stored in CDIS and RIS. AMR traceability to other scientific documents and models are supported both in TDMS and DIRS. The AMRs consist of data, analyses, models, software, and supporting documentation that will be used to defend the modeling effort for evaluating the post closure performance of the potential Yucca Mountain repository system.

- **Abstracted Models:** ***Develop nominal and disruptive event scenarios for retained FEPs.*** This stage of the pyramid involves the development of abstracted models. These abstractions are progressive simplifications of the conceptual models of physical and chemical processes to more compact, useable numerical models. The numerical models include mathematical representations or abstractions of the conceptual models of the FEPs and/or scenarios that contribute to overall repository performance. The mathematical models consist of quantitative expressions of the process models such that they can be used together to simulate repository performance. The mathematical models might include algebraic expressions, ordinary differential, partial differential, or integral equations characterizing accepted conservation laws (e.g., conservation of mass, energy, or momentum), and appropriate constitutive equations that describe material behavior in the domain of the conceptual model.

  An example of one of the process models abstracted in mathematical and numerical form is a model describing the flow of water infiltrating at the surface of the land and then percolating through the unsaturated zone (UZ) above the water table. Such a model would incorporate equations describing fluid flow and probable fluid interactions between the rock matrix and fractures in the rock as well as descriptions of any other hydrologic, physical, and chemical processes needed to describe how water flows throughout the rock mass of the UZ.

- **Total System Models:** ***Implement nominal and disruptive event scenarios.*** The top level of the pyramid consists of the integrated total system model. The total system model is a numerical model used to simulate the behavior of the Yucca Mountain repository system. This total system model, the basis for the TSPA-LA Model, incorporates the abstracted process models and/or the AMRs that describe the model components and their sub-models, from their development to their implementation.

4.0 TDM SYSTEMS BUSINESS OBJECTIVES

The scientific investigation process (Figure 2) supports TSPA for the Yucca Mountain license application and is used to identify, create, authenticate, and maintain technical data and technical products. TDM Systems is primarily structured to maintain the integrity, traceability, and pedigree of technical data. TDM Systems, which supports SIP, is organized into two main functional areas, the Technical Data Management System (TDMS) and the Document Input Reference System (DIRS).

TDMS, along with an array of other applications with which it interfaces (e.g., CITIS, AFS, SCM, Lotus Notes, RIS, TIC, and CDIS), supports the scientific investigation process areas for (1) the input of technical data generated by authors, (2) the creation and input of metadata and index information associated with the technical data in order to maintain pedigree, and (3) the search and access by authorized users to the information contained in the TDMS database.

DIRS, also in conjunction with other applications (e.g., RIS, CDIS, Infoworks, TIC, and Lotus Notes), supports the scientific investigation process to meet three management goals: (1) ensure the existence and correct formatting of bibliographies and works cited in program documents; (2) assign cross-references to allow tracking and resolving references originally designated as "*To Be Verified*" (TBV); and (3) cross-reference document inputs and document products.

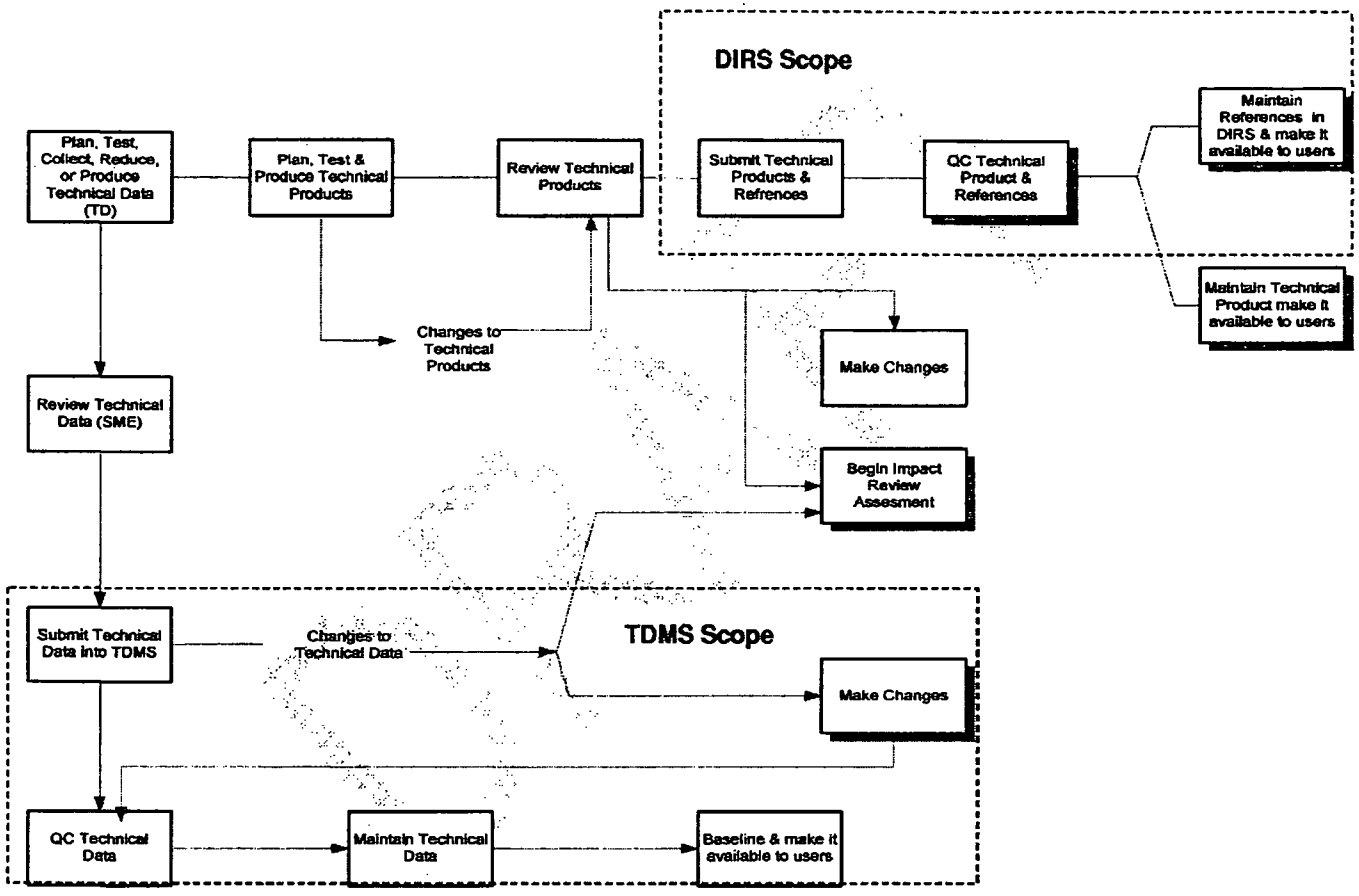


Figure 2. Overviof the Business Function: Scientific Investigation Process

4.1 Current TDMS

The TDMS is currently organized around three main functions (i.e., data entry, data quality assurance, and data retrieval/usage). These involve five separate sets of roles and responsibilities (i.e., originator, reviewer, data coordinator, database administrator, and records coordinator) as depicted in [redacted] in Appendix A.

Data Submittal: In order to develop the Total System Models that are used to simulate the behavior of the repository (discussed in and Section 3.0), scientists and engineers use collected, reduced, and/or produced data. These data come in various formats and categories (e.g., meteorological, geological, physical measurements, etc.). TDMS is used to submit new data, changes to these data, and information/attributes about these data. (See [redacted] in Appendix A for data attributes that are entered.) This activity is currently accomplished in coordination with the data coordinator and the originators at various OCRWM-wide organizations and/or agencies. (See [redacted] in Appendix A for OCRWM-wide organizations and agencies that submit or use TDMS.)

Data Quality Assurance: After the data and its associated attributes are entered into TDMS, a subject matter expert (SME) reviews the data. This process is currently a manual process and is done outside of TDMS. Once reviewed by the SME, the TDMS database administrator (DBA) performs a quality assurance (QA) check on both the data and associated attributes to ensure all attributes are accurate, complete, and comply with data integrity rules. Any discrepancies are referred to the originator for resolution. Upon completion of this QA check, the data is locked and no further manipulation is permitted.

Data Retrieval/Usage: After being locked, data are published as records and made available to the engineering and scientific community in organizations and agencies across OCRWM.

4.2 Current DIRS

DIRS is currently organized around three main functions (i.e., reference entry, reference verification, and reference usage) that involve three separate sets of roles and responsibilities (i.e., originator, reference locator, and DIRS administrator). [redacted]a depicts the DIRS overall flow; then sub-processes in this flow are shown in more detail in [redacted] (both in Appendix A): Verify Reference Process and Figure 6c: DIRS TBV Resolution Process.

Reference Entry: DIRS allows authors/originators at the various OCRWM-wide organizations/agencies to (1) submit references and associated attributes that are cited in the work products of the authors/originators if those references and associated attributes do not already exist in DIRS, and (2) track verification of cited references. (See [redacted] for the attributes that are entered.)

Verification: After a reference has been submitted, the DIRS Administrator (1) verifies the entered information to ensure the existence and correct formatting of bibliographies and work cited, and (2) assigns cross-references to allow tracking and resolving references originally designated as TBV.

Usage: Once a reference has been located and verified, it can be accessed and cited as an already verified reference by scientists and engineers in their work products at the various OCRWM-wide agencies and organizations.

5.0 INFORMATION SYSTEMS ARCHITECTURE

5.1 Current Information Technology System

A collection of five major databases (ATDT, MW, SPA, TDP, and GI) are collectively referred to as TDMS. These five databases, user interface screens, and processes require extensive and complicated manual manipulation. Figure 8 depicts the context of TDMS. Processes in each technical area are governed by seven specific desktop procedures that detail the step-by-step mechanics and control of input into each of the five databases. These processes are reliant upon strict adherence to procedures as well as the manual accuracy and consistency of TDMS personnel for success. Additionally, most operating system software, middleware, database management system software, and programming languages are outdated technologies. The TDMS interfaces/links (manually or via HTML) to numerous other outside systems (i.e., CITIS, AFS, SCM, Lotus Notes, RIS, TIC, and CDIS) so that TDMS can support SIP. (Refer to Figure 2 in Appendix A for the business function.)

DIRS provides HTML links to the ATDT in addition to outside systems (i.e., RIS, CDIS, Infoworks, TIC, and Lotus Notes). DIRS is governed by three specific desktop processes to which the DIRS Administrator must strictly adhere for success.

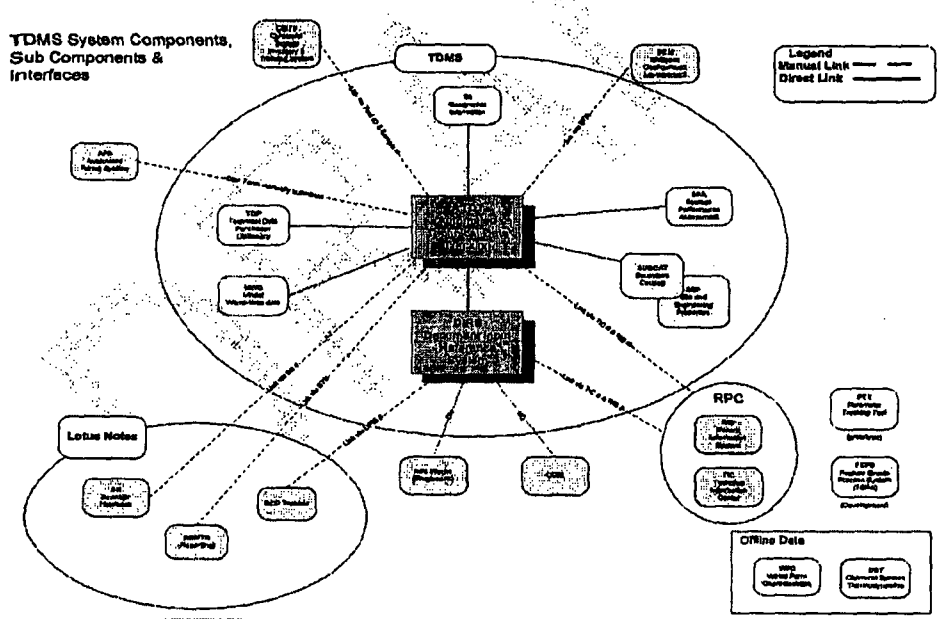


Figure 8. *Overview of Technology: IT System Diagram ("As-Is-State")*

On the following page, **Error! Reference source not found.** provides descriptions of the seven components represented in Figure 8 that primarily support TDMS, including the purpose, major capabilities, actors, and compliance documents for each.

The flow of data through the current TDMS is shown in Figure 9 and the flow through DIRS is shown in Figure 10.

Table 1. Description of Systems within the Scope of TDMS

Systems	Purpose	Major Capabilities	Actors	Compliance Documents
ATDT	Commonly referred to as the "card catalog;" a master indexing system that provides traceability from the highest level data developments and interpretations back to the original source data using Data Tracking Numbers	• Trace data to source data • Trace data to analysis model	• Originator • Data coordinator	• AP-SIII.3Q • LP-SIII.2Q-BSC
GI	Contains spatial data providing location information for test activities, facilities, roads, and geo-science features within the Yucca Mountain study area	• Maintains traceability to the geographic data in SEP		
SPA	Used to compile performance assessment modeling input files. Files containing graphical and tabular data are used to facilitate analytical evaluations of the site and repository system design.	• Submittal of performance assessment modeling input data • Maintain input data • Report	• Originator • Data coordinator	• AP-SIII.3Q
SUBCAT and SUBCAT/SEP	SUBCAT is designed to work in conjunction with the SEP database. It contains site investigation and laboratory test data as well as engineering analysis input data organized by Parameters or a Specific Data Tracking Number.	• Submittal of scientific data • Retrieval of scientific data • Maintains a catalog of technical data • Maintain verification status of the submitted scientific data	• Originator • SEP Database Administrator • Processing Analyst • Verification Analyst	
MWD	Comprised of input data and output results of the various site evaluation models, including process and intermediate level analytical models	• Submittal of scientific data • Retrieval of entire or partial models		• AP-SIII.3Q
TDP	Dictionary of standardized technical terms that is used to identify all YMP technical data. These standardized terms (i.e., parameters and attributes) provide the common language needed for storing and retrieving the project technical data generated by or required for site characterization activities, socioeconomic and environmental assessments, design and development activities, and performance assessment.	• Maintain parameters and attributes • Parameter validation • Attributes validation • Reports	• Originator • Data administrator	• AP-SIII.3Q

Yucca Mountain Project: TDMS DRAFT Concept of Operations Version 1.0

Systems	Purpose	Major Capabilities	Actors	Compliance Documents
DIRS	Used to capture, track, and, maintain traceability for references in major YMP legally required, government-deliverable, and publicly accessible documents	• Access to reference information • Submit reference • Cited references verification • Reports and Queries	• DIRS Administrator • TBV Administrator • Author • Researchers	• LP-3.15Q-BSC • LP-3.21Q-BSC

5.2 Technical Architecture

Performance Characteristics: Error! Reference source not found. provides a list of the performance characteristics in the existing systems.

Table 2. Performance Characteristics of Current System

Characteristic	TDMS Systems						
	DIRS	ATDT	SEP (SUBCAT)	NAS	MWD	SPA	GI
Interface	Web	Terminal and Web	Web	Terminal and Web	Terminal and Web	Terminal and Web	Terminal and Web
Volume/data set	< 10KB	< 10KB	varies	1K – 500GB	varies	varies	varies
Frequency of Data Additions	daily	daily	weekly	weekly	infrequent	monthly	infrequent
Frequency of Data Updates	daily	daily	rarely	daily	rarely	rare	rare

Quality Attributes: Data is stored in its native file format. During submission, all related files are compressed into a single zip file that is submitted and stored on the NAS.

Sustainment: Some capabilities, although tenuous, are in place for sustaining the current system.

- User-interfaces are web-based applications and terminal emulations via Ingres Application by Forms (ABF) that display as 80 x 24 character windows. Some interface programs are written for specific software and platform languages, but are not portable.
- Maintenance is performed either in-house or contracted out to third-party sources.
- The system is operational during normal business hours. Batch processing is scheduled and run at night to move data and backup files. During this time, user access is restricted to system personnel.
- BSC was the YMP Lead Laboratory for many years and much of the data, procedures, networks, etc. are owned and/or maintained by BSC.

Security and Emergency Operations: Limited security and emergency operations procedures are in place:

- The servers and applications are shared resources on the Yucca Mountain system. Each requires a login username and password for authentication and establishment of user rights.
- The design of the existing user interface requires web-based authentication. The existing system requires that the web application have full rights to the databases.
- Operational procedures are in place to check the accuracy and validity of reference items. These procedures are outlined in a series of user manuals for each subsystem. There are no known referential integrity constraints built into the system to automate the validity of references.

Internal Components: Error! Reference source not found. provides a list of the internal components of the current system.

Table 3. Existing System Internal Components

System	Application Architecture	Application O/S	DMBS and Storage	DBMS HW and OS	Middle Tier
ATDT	VMS terminal emulation. For data entry Web Based Queries using CGI and C code. FTP and HTML templates for creating scripts for FTP. Ingres ABF. URL links to external systems – RIS, DIRS, GENTS, SCM, CDIS	HP Open VMS Web 5.4.0 ABF – R01V15	Ingres 6.4 for metadata NAS storage for files in native format 1GB DB storage 4TB storage available on NAS	HP Open VMS for Ingres. Windows File sharing for dataset files	Apache 1.3.26
SEP	CGI and C code FTP. Interface with SUBCAT	6.1.0	Open Ingres 1.2, Ingres 6.4 7GB DB storage 4TB storage available on NAS	Solaris 2.6 HP Open VMS Windows File sharing for dataset files.	Apache 1.3.26
SUBCAT	OpenRoad 3.5 UNIX CGI and C code. FTP. Interface with SEP	6.1	Open Ingres 1.2	Solaris 2.6	None
DIRS	Web based using CGI and C code. URL links to external systems – RIS, RCD tracker, TIC, ATDT, CDIS	2.3.0	Ingres 6.4	HP Open VMS for Ingres	Apache 1.3.26
GI	Query By Form (QBF). Data feed one at a time from ArchInfo. URL links to ATDT, SEP	Web 3.2.0	Ingres 6.4	HP Open VMS for Ingres	Apache 1.3.26
MWD	Web based using CGI and C code. Partial HTML file describes the SUBCAT data in detail	Web 3.2.0	Ingres 6.4	HP Open VMS for Ingres	Apache 1.3.26
SPA	Web based using CGI and C code. Partial HTML file describes the SUBCAT data in detail	Web 3.2.0	Ingres 6.4	HP Open VMS for Ingres	Apache 1.3.26
TDP	Web based using CGI and C code. Web is read-only. Few persons are allowed to update via ABF interface. Data is stored in a table in the ATDT database.	Web V2.1.0 ABF 3.0.0	Ingres 6.4	HP Open VMS for Ingres	Apache 1.3.26

System Network: Key elements of the BSC network and its maintenance are listed below:

- The principle servers are Hewlett Packard (HP) Alphas running on open VMS. The Network Area Storage, which is also on the network, stores a large portion of the data sets in their

native file formats. The Data Base Management System (DBMS) is Ingres 6.4, which is no longer supported, and Open Ingres 1.2, which is a more recent open source version.

- The systems are supported as-is with no plans for hardware upgrades or software enhancements until a replacement system is designed and implemented. Communication between systems is through TCP/IP on the local BSC network. Hyperlinks allow access to files internally on the network and to web-accessible files on the Internet.
- The SEP database subsystem was stored in data sets in Ingres tables prior to 2002. Since 2002, data submissions are stored in their native format on the Network Area Storage (NAS). Some information about the data sets is still stored in the database. The largest data package submitted to date was ~500GB in a single zip file.
- Services are delivered via a web interface using Microsoft Internet Explorer 4.0 as the minimum. The Ingres Application By Form (ABF) is a native UNIX process that runs on the HP Open VMS through Microsoft's terminal emulation program. Data entry and display screens are shown in 80 characters by 24 lines.
- MWD, SPA, and GI are HTML pages with HTML links to ATDT. There are file server links to data files on the NAS for native data files. When links are selected, a separate web page is opened with the data from the database or file. Data files are uploaded in native format via ftp to the NAS system.
- User desktop computers run on Windows 2000 with a web browser and terminal emulation program. Client-side software for Ingres must be installed to run applications written in Ingres 4GL.

6.0 CURRENT SYSTEM GAP ANALYSIS, DESIRED CHANGES, AND RECOMMENDATIONS

TDMS and DIRS were developed to manage technical data and work products for the repository to (1) ensure the integrity and quality of this data, and (2) maintain traceability for references to this data complying with the Yucca Mountain TSPA. However, the analysis of the current system identified significant gaps, issues, and risks in meeting these objectives. The primary issues and risks are identified below with a detailed listing provided in Appendix F.

The new system must track data items systematically through the system from end-to-end. Conclusions developed and published for the Licensing System must be able to verify how data was developed throughout the analysis and modeling processes. Referential integrity must be maintained by the database system to ensure the consistency and accuracy of the data. The following TDM Systems changes are required to address the issues and risks identified in this section. Appendix F provides a more detailed view of these changes.

6.1 Business Process

Issue: Suboptimal Business Process

The following examples provide a view of the current Scientific Investigation Process gaps and weaknesses. They identify some of the areas not addressed by the current process:

- Vendor data, also known as Qualified Supply List (QSL) data are stored in TIC. When this QSL data is changed, there is no provision for updating the changes into TIC. Additionally, there is not an IRAN process in place to notify the users of the old version that they need to do an impact analysis
- Data Administration is allowed to change the data into TDMS before the receipt of replies to IRAN. Also, receipt of IRAN replies has an open time limit.
- The review of originator-data-submitted packages first by the data coordinator and then again by a reviewer might be unnecessarily redundant in the current process. (Perhaps it is currently necessary because so much of the process is manual and it helps catch errors.)
- The document owner list is not always current. When the original owner leaves the organization, there is no process in place that updates the document's owner name.
- USGS data is allowed to be directly submitted into RPC instead of via TDMS. This undermines data quality since changes and qualifications, among other things, are not controlled in RPC as they would be in TDMS.

Desired Change and Recommendation: Optimal Business Process

Create a streamlined optimal business process in order to gain a common understanding among various organizations and agencies that implement specific processes and get rid of duplicated effort while removing inadequacies in the existing process.

- *Enterprise Business Modeling and Value Stream Analysis* are recommended in order to identify business areas that are either not addressed or are weak. This approach will also help the IT Integration team to target and prioritize business areas that need automation. Individual projects can then be evaluated with an understanding of how their effort fits into the overall business.
- *"To-Be-State" for TDMS Process has been redesigned. See [redacted] in Appendix A.*
- *"To-Be-State" for DIRS (TBV Process) should be redesigned during the next phase.*

6.2 Process Areas

Issue: Unsupported Process Areas

Error! Reference source not found. maps the business processes used for scientific investigation to the information systems that support those process areas. Mapping the two provides a comprehensive picture of possible gaps. Parts of the business processes (see [redacted]) are supported by TDMS, DIRS, and other peripheral systems while critical processes are executed manually since the processes are not currently supported by any information system(s). For example:

- When QC of the technical data is initiated, documented evidence of a completed technical review for Un-Q data is provided manually as a hardcopy sent to the TDMS DBA. That hardcopy is then scanned into and maintained by RIS. This causes delays in accurately identifying data qualification, creates unnecessary manual labor causing bottlenecks for the data coordinator and the originator, and is prone to error.

- The Impact Review Assessment Notification Process is supported manually causing errors and delays. This has a direct impact on data Integrity and traceability.

Table 4. Mapping of Business Processes to Information Systems

Business Processes	Information Systems												
	SCM	TSPA	ATDT	GI	SEP	SPA	TDP	MWD	CSITS	RIS	CDIS	TIC	DIRS
Screened FEPS maintained		●	▼	▼	▼	▼	▼	▼					
Plan test and Create Technical Products (i.e., AMR, SNB)	N/A												
Review Technical Products	N/A												
Submit Technical Products	Manual												
Submit Product Reference	Manual												
QC Product and Product References	Manual												
Maintain references and make available to users													●
Maintain technical product and make available to users										●	●	●	
Plan, collect, reduce technical data	N/A												
Review Technical Data (SME)	Manual												
Submit Technical Data	Manual												
QC Technical Data	Manual												

Business Processes	Information Systems												
	SCM	TSPA	ATDT	GI	SEP	SPA	TDP	MWD	CSITS	RIS	CDIS	TIC	DIRS
Trace developed data to source data	Manual												
Maintain Technical Data's attributes and make available to users			o	o	o	o	o	o	o	o	o	o	
Impact review assessment	Manual												

Note: Blue highlighted columns indicate systems within scope of this project
N/A represents activities that take place in originator organization OCRWM wide and various tools used with them
o represent partial support

Table 5. (continued)

Desired Change and Recommendation: Supported Process Areas

Some areas such as the Impact Resolution Assessment should be automated to the greatest degree possible. However, no definite recommendation can be given as to whether currently unsupported areas (see Table 4) should be supported further. Solid understanding and then re-engineering of the complete Scientific Investigation Process is necessary before such a recommendation can be made. Enterprise Business Modeling and value stream analysis will provide specific business missions that need to be supported; systems can then be developed with that enterprise view in mind.

6.3 Technology

Issue: Dated Technologies

Most of TDMS operating system software, middleware, database management system software, and programming languages are outdated and often unsupported technologies on the BSC network.

- Most systems are supported as-is with no plans for upgrades or software enhancements until a replacement system is designed and implemented. Maintenance and support are extremely limited or non-existent. The Ingres data base management system is so outdated, it is no longer supported by the vendor. This puts the system at high risk if a major hardware failure or software defect in the application or database were to occur. Catastrophic failure of the hardware would have a severe impact on the ability to continue the mission of TDMS.
- The physical architecture also limits the expandability of the system. Newer technologies are not compatible with the older software.
- Some of the software, especially Ingres, also has severe limitations in accommodating ever-increasing data sets. Previous developers have overcome these limitations by extending logical records across several physical tables within the database, but this solution presents additional risk of error and of failure.

- Client-side software for Ingres must be installed to run applications written in Ingres 4GL.
- Some interface programs are platform specific and therefore are not portable.
- The SEP database subsystem is stored in data sets in Ingres tables that pre-date 2002. Since 2002, data submissions are stored in their native formats on the NAS. However, some information about the data sets is still stored in the SEP database. The largest data package submitted to date was ~500GB in a single zip file.
- Processes in each technical area are governed by ten specific desktop procedures that detail the step-by-step mechanics to setup and control user interfaces, screens, and each of the five databases. Currently TDM staff employs error-prone, low-level computer commands and techniques normally used by technical computer database administrators.

Desired Change and Recommendation: Updated Technologies

It is recommended that longstanding outdated TDMS and DIRS be overhauled.

- Old codes and infrastructure with very little documentation that are hard to maintain should be replaced with new standards.
- Standardizing will improve agility, cut costs, and accelerate responsiveness. Stovepipe information system applications developed and deployed on disparate platforms, can be integrated into a common, seamless user environment.
- The redevelopment and upgrading of the current system will cause data packages to upload to network storage areas with minimal user interface. The system will automatically maintain directory structures and files.
- The new system must be developed with a standard tool set.
- The presentation and user interfaces must be common across subsystems. Data entry processes must be standardized across all subsystems.
- Internal tracking methods must be employed as data moves through the system. A data-centric approach must be utilized.

6.4 Manual Manipulation

Issue: Extensive Manual Manipulations

Although current functions may guarantee that correct processes are followed, TDMS cannot guarantee the "correctness" or authenticity of the data. The data pedigree is suspect. In order to accomplish many of the operational procedures, manual intervention is required, making the system time consuming, error ridden, and labor intensive. For example:

- Operational procedures in place to check the accuracy and validity of reference items and data are outlined in a series of user manuals for each system. There are no known referential integrity constraints built into the system to automate the validity of references and the system lacks implemented controls (i.e., business rules) and verification checks. Data is manually reviewed and retrieved by quality control personnel to validate and verify

information. This review process is used for both the technical data entered into TDMS and documentation entered into DIRS.

- Manually tracing developed data to its applicable source data (SNB's, AMR, its page numbers, etc.) causes inconsistencies in the data pedigree.
- Manually assigning DTNs for data and scientific information causes inconsistencies and missing DTNs, resulting in major problems in finding the accurate, timely, and consistent information in TDMS that is required to support LA.
- Originators currently create hard copies of submittal packages that are first sent to data coordinators, then reviewers, and finally to a DBA who actually submits the data into TDMS. This manual transfer of the submittal package is tedious and redundant causing bottlenecks and unnecessary work when time could be spent supporting other important activities.

Desired Change and Recommendation: Automation

The desired state should incorporate business rules in computer system code. The ***Business Rules to be incorporated have been identified and documented in TDMS and the DIRS Requirements Document.***

- Validation and verification must be integrated into the system eliminating manual review and re- review for quality control.
- An automated workflow process will allow each functional role to review and approve data as it flows through the system.
- System audits will verify and track the approval of items submitted and processed within the system.
- The process and rules by which the flow of documents, information, and tasks are passed from one participant to another have been documented in the Requirements Document.
- It is recommended that DIRS's TBV process be re-designed during the next phase.
- Data and Text Mining techniques should be used to shift through terabytes of unstructured data in eliminating duplicate data, resulting in better traceability and a better search capability. (For complete analysis of DATM see [redacted].)
- In the current system, DTNs are used for traceability, unique identification, and searches; this will instead be supported by the system as change management functions, thus eliminating the need for creating and tracking DTNs.

6.5 Application Architecture

Issue: Stove Pipe Applications

Each of these functional areas has supporting applications operating in a legacy infrastructure environment consisting of "stove pipe" systems and data supporting unique business functions. Stove Pipe applications not designed to support reusable code are expensive to maintain. Lack of automated interfaces between them greatly introduces erroneous data into the system due to multiple entry points. Critical information must be copied and pasted from one application to

another. Figure 8 provides a comprehensive look at peripheral stovepipe applications currently providing data (manually) to TDMS and DIRS. These system interactions are critical for accomplishing TSPA. In addition to data sharing, there are other reasons for integration. For example:

- Tracking open and closed statuses of documents and data and other workflow management is accomplished through Lotus Notes, which is by and large manually integrated to TDMS.
- Integration between the TSPA System which maintains FEPS and TDMS is crucial for complete TSPA traceability.
- Integration between SCM and TDMS is necessary for users to have access to the latest version of the software used to create models in the Model Warehouse.
- There is no integration between CDIS and DIRS. CDIS does not know when a document in DIRS is locked, and DIRS does not know of changes in a document's status in CDIS (i.e., a document could be cancelled in CDIS).

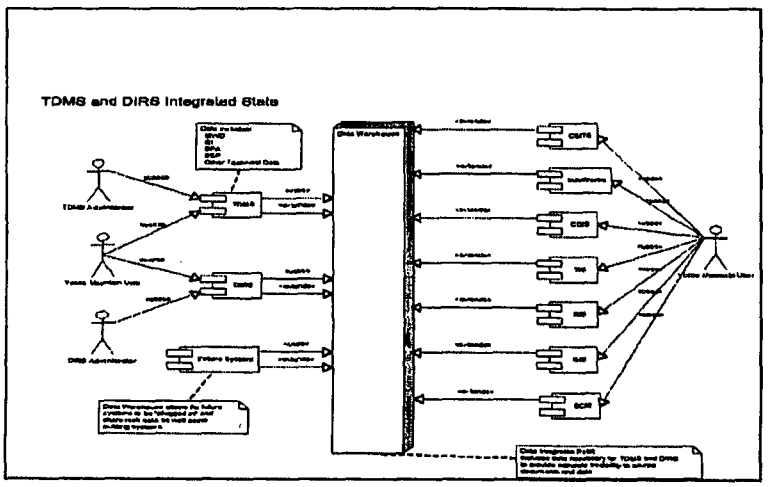


Figure 12. TDMS and DIRS Integrated State ("To-Be-State")

Desired Change and Recommendation: System Integration

A Data Warehouse is recommended for System Integration (See Figure 12 above for TDMS DIRS integrated state on the previous page and see Appendix A for a complete data warehouse analysis). The many benefits to building and owning a data warehouse can address many of the issues and gaps that have been identified with the YMP TDM Systems (see Appendix B). A data warehouse is a subject-oriented database designed with enterprise-wide access in mind. It

provides tools to satisfy the information needs at all organizational levels—not just for complex data queries, but as a general facility for getting rapid, accurate, and often insightful information. It is designed so that users can recognize which information they want and access that information using simple tools. Data in a data warehouse differ from operational systems data in that they can only be read; they cannot be modified. Operational systems create, update, and delete production data that "feed" the data warehouse. The principal reason for developing a data warehouse is to integrate operational data from various sources into a single and consistent repository that supports analysis and decision making within the enterprise.

- **Integrating disparate data into a single repository.** The data that currently resides in various disparate systems outside of TDMS, such as Lotus Notes, CDIS, RIS, TIC, SCM, and CSITS, can be integrated into a central repository where it is more easily accessible for TDMS subcomponents, such as ATDT and DIRS, to capture, track, and maintain traceability for references in YMP documents.

Formatted: Bulleted + Level: 1 + Aligned at: 0" + Tab after: 0.25" + Indent at: 0.25", Tabs: -0.13", List tab + Not at 0.25"

- **Reducing manual processes.** With a data warehouse in place, the manual processes that are currently being used to track pieces of information in TDMS specific systems, along with those outside of TDMS, will be reduced substantially since information can now be obtained from a central data repository.

Formatted: Bulleted + Level: 1 + Aligned at: 0" + Tab after: 0.25" + Indent at: 0.25", Tabs: 0", List tab + Not at 0.25"

- **Enhancing data quality and integrity.** The integrity and quality of YMP essential scientific and engineering data would be enhanced since part of the warehouse building process includes checks to detect inconsistent and duplicate data. As a result, the data must be cleaned up in the source systems before being loaded into a data warehouse.
- **Enabling queries that cut across different databases.** The need to have stronger mechanisms within TDMS to pull data from various data sources can be addressed via the data warehouse. Through a central data repository, queries and custom reports could be easily built and maintained in data warehouses, thus decreasing the workload on transactional systems.
- **Providing a platform for analytical queries and integrated reporting.** A data warehouse repository integrates the data from all of the associated YMP systems to provide users with a platform upon which querying, reporting, and analytical tools can be based. For example, the ability to obtain and report on the qualification status of a DTN can be easily attained through a data warehouse.

6.6 Security and Maintenance

Issue: Suboptimal Security and Maintenance

- There are security issues and risks, due in part to the use of outdated technologies, practices, and procedures.
- The design of the system makes it necessary for TDMS administrators to have full access to the file server and production database so that they can publish the static web pages, upload datasets, and update the database when they receive new or changed datasets. Because of this, administrators have the ability to intentionally or accidentally manipulate

production data without gong through the application, thus bypassing access controls and any possible audit.

- Some user-interfaces allow the user to perform searches of TDMS. The user can select all or part of the indexed information to search. The index information is stored primarily in the ATDT database.
- There is no off-site backup facility in place to facilitate emergency operations should a catastrophic event occur.
- Some original data sources, especially copyrighted material, are stored in hard copy format. In many cases, only one copy of the source exists. This makes it difficult to reference this material and puts the reliability and traceability of such references at risk.
- Some data exist in Waste Form Characteristics and Chemical Species Thermodynamics that are now off-line. The Yucca Mountain staff believes that all relevant data in these systems have been migrated to TDMS, but this cannot be confirmed.

Desired Change and Recommendation: Secure and Backed-up System

Users must be allowed to add or modify data based on their job functions or roles. Each subsystem must create and maintain a list of users and their rights. Privileges must be dependent on the role a user is assigned for each subsystem. Access and permissions must be limited to the lowest level required. These range from read-only to full update capabilities. Auditing procedures must be implemented to track data changes.

User access controls using roles/groups can be implemented in the "to-be state."

7.0 CHANGES CONSIDERED BUT NOT INCLUDED

Consolidation of TDMS and DIRS was considered but then dismissed because mapping of the functional areas (see Table4) supported by DIRS and TDMS does not show an overlap between the two systems. They support unique business functions: DIRS supports document references whereas TDMS supports data management. Consolidation of the two systems would only make sense if all other peripheral systems (e.g., RIS, TIC, and CDIS,) that support SIP were also included. In order to do that, further analysis of the peripheral system and re-engineering of business processes with the objective of making the process more efficient has to occur first. Also, because a single consolidation implementation must support the needs of many agencies (OCRWM wide), it must incorporate agility, flexibility, compatibility, and ease of use to ensure that it is adapted quickly, seamlessly, and with few, if any, service disruptions. We do not see this as a possibility by implementing one monolithic system in the current environment.

8.0 REUSE/BUY/MAKE ANALYSIS

Current data management and tracking capabilities were identified and surveyed to determine general available functionality. Capabilities include reuse, COTS, and options

developed as in-house solutions. High-level gap analysis was performed to filter out obviously incompatible solutions.

8.1 REUSE ANALYSIS

The following Sandia National Laboratories Tools could potentially contribute to the TDMS Development Effort:

DART SDM TOOL SET (Design Through Analysis Team Simulation Data Management Tool Set)

The DART project encompasses a number of tools and technologies (see [redacted] for a complete analysis). Not all DART offerings are relevant to the Yucca Mountain Project (YMP) Technical Data Management System (TDMS). The portion of the DART toolset that provides document and data management services is called the DART Simulation Data Management Toolset, or DART SDM Tools. It is the DART SDM Toolset that is most likely to be of interest to a TDMS redesign.

One of the primary strengths of the DART environment is its data management model. Included in this model are several features of potential interest to TDMS:

- A mechanism to group data into logical sets called artifacts, and to organize artifacts into a structured project
- The ability to keep a history of data changes, including attributes such as who made a change and when
- The ability to annotate data with time- and user-stamped comments
- The ability to track input-to-output dependencies between artifacts, including which specific versions of artifacts are involved in the relationship
- The ability to restrict access to data to specific authenticated users, groups, or roles
- The ability to extend the existing data model to accommodate new types of data

The DART data model is implemented within the DART SDM Toolset. A more complete description of the DART data model can be found in *An Introductory Guide to the DART Environment*, currently available as Web FileShare document 448860.

AMR Exploration Tool

The AMR Exploration Tool is designed to facilitate:

- Identification of data to be used in the defense of the License Application for Yucca Mountain
- Determination of the qualification status of this data through traversal of the multi-faceted interrelationships among this data

The tool addresses this need by providing a means to visualize the graphical relationship between data "nodes," where each node represents a particular piece of data (e.g., DTN, Record, and CDIS documents). The tool provides user-directed navigation of this graph to perform data traceability for qualification purposes. Navigation can occur either through following explicit links in the node records or by choosing from a list of recommended node neighbors automatically generated based on a textual query from the user. A record of

the trace performed, trimmed by the user to include just those elements that were determined to be germane, can be saved and re-accessed later by the user. In addition, the tool provides a means of sharing trace reports with others in the form of self contained files or hard copies.

Issue Management Database

The Issue Management Database helps users collect, refine, query, and track issues, including those related to risk and vulnerability on a basis of collaboration. This flexible web-based application, available on Sandia National Laboratories' Secure Open Network (SON), can define and organize an unlimited number of issues and issue types as well as their associated risks, mitigation plans, and responsible owners. The integrity of this database is secured using information access controls. It has the following features:

- Module-based issue management and access control
- Unlimited, fully configurable modules, issue types, data attributes, and attribute options
- Storable user-defined queries of both personal and public scopes
- Free text queries
- Advanced querying tools
- Export of query results to Microsoft Excel files or tab-delimited text files
- Import and export of issues through XML
- E-mail notifications
- A facility to attach comments and electronic files to issues
- Detailed, automatically maintained historical information for each issue
- Extensive customization through an administrative user interface

Formatted: Bulleted + Level: 1 + Aligned at: 0.38" + Tab after: 0.63" + Indent at: 0.63", Tabs: 0.25", List tab

Formatted: Indent: Left: 0.38", Bulleted + Level: 1 + Aligned at: 0.38" + Tab after: 0.63" + Indent at: 0.63", Tabs: 0.25", List tab

Formatted: Bulleted + Level: 1 + Aligned at: 0.38" + Tab after: 0.63" + Indent at: 0.63", Tabs: 0.25", List tab

SDDB (The Stockpile Dismantlement Database)

- SDDB manages all data supporting an environmentally conscious disposition and the safe handling of hazardous materials of all weapons hardware. The SDDB system is currently organized around three main functions: (1) Data Entry, (2) Data Quality Assurance/Quality Control (QA/QC), and (3) Data Retrieval. SDDB is designed for use by various personnel: Waste Management, Shipping, Auditors, Accident Response Groups, Industrial Hygiene, Safety Planners, Design Engineers, Component Engineers, and Maintenance Personnel

- Allows electronic transfer of data to DoD users via scheduled data submissions to DTRA (Defense Threat Reduction Agency)

- Design ensures information security through trained staff, password control, limited networks, and read-only privileges

- Graphical displays of the weapons hardware disassembled for the purposes of safe disposal, recycling, treatment, or reuse as the disposition tree levels up and down, and interchangeability of parts. This feature also allows the user to identify the

characterization status of each part in a disposition tree. The use of various colors signifies the various statuses of part characterization and QA/QC (i.e., white represents not started, blue represents started, green completed, etc.) Graphical examples are shown in Appendix B.

•

8.2 COTS ANALYSIS

The following commercial off-the-shelf (COTS) tools were surveyed and analyzed as potential solutions. (For a complete COTS analysis, see Appendix E.)

WebPE

WebPE was presented to a segment of BSC last year as a possible replacement for TDMS. WebPE, Inc. is described as being a specialist in environmental software and portal technology. Its listed capabilities include:

- Workflow support such as version control, approval process (parallel of series), event management, and alerts.
- "Need to know" access and deployment for large, complex groups
- Dynamic reports providing instant, relevant, and accurate results
- Includes both documents and data
- Portal integration
- Tools to store and technical libraries to work with discrete data including:
  - Analytical Methods
  - Chemistry
  - Field Measurements
  - Geography
  - Hydrogeology
  - References
  - Sampling
  - Thresholds/Criteria
  - Well Construction

Formatted: Bulleted + Level: 1 + Aligned at: 0.38" + Tab after: 0.63" + Indent at: 0.63", Tabs: 0.25", List tab

WebPE handles the creation, input, storage, and management of structured data. It also can store and manage unstructured data. A reference component exists to link multiple references to a document, but it does not appear to support the DIRS process as-is (in relation to TBVs specifically). Application Software Platform is based on the Oracle Portal and the Oracle Application Server. The database platform is Oracle 10g. Supported OSs include Linux, Sun, and Windows. Licensing is $17,000 per CPU with no further client, group, or user license fees. Upon initial investigation, WebPE appears to be the closest fit of all the COTS products evaluated, but there are concerns that should be addressed.

- The company is in its infancy and has been in business for little over a year, which could lead to longevity issues with the company, potentially leaving little or no product support.
- The product's structured data input and creation components would most likely not be adopted as engineers and scientists have their own tools (which they are familiar with) for generating this data.

- The document management and workflow component appears to be the most relevant component of the product in relation to TDMS. However, this is only a portion of the product's capabilities, meaning this component may take a back seat to the product's other capabilities.

WebPE may be able to replace one or more systems related to TDMS, but further analysis should be performed to determine whether the replacement would provide significant benefits over the current system. If this tool is considered or selected as a complete or partial solution, at the minimum, a demo version should be tested first.

SharePoint

Microsoft's SharePoint appears to include an Enterprise Content Management (ECM) component. Its capabilities include:

- Document Management: check in and checkout, versioning, content types, metadata
- Workflow: review and approval, light-weight process customization
- Policy and Compliance: records management, expiration, IRM protection, auditing

SharePoint now includes workflow capabilities built on a Windows Workflow Foundation that are integrated with the document management repository and forms capabilities. SharePoint is currently slated to be used on the YMP network and should be considered as a possible solution. Pros include optimum traceability, tracking, and data integrity. Cons include the highest cost and the greatest data migration efforts.

Documentum

Documentum is a leading ECM product owned by EMC software. It has the ability to record the complete history of a document's movement and custody, including the daily activities of creating, transferring, checking-in and -out, and destroying documents. It can support inventory and audit reporting needs, has bar-code capability, and provides security at the functional level. The strength and focus of the Documentum products are for managing electronic documents; however, it is capable of managing paper documents as well. Sandia has purchased a 10,000 user license for the records management and content management components to be used for the Electronic Archiving project with the Records Management group. Further investigation would be required to verify whether this licensing would allow the use of Documentum to meet the needs of classified document management.

Stellent

Stellent is a web-based content management product. As with most current ECM products, its focus is on and its strength is in managing access to electronic documents. It has good capabilities in tracking electronic access to documents, but has significant shortcomings in tracking the movement of physical documents. Although intended as a version control mechanism, it does have check-in and -out capabilities that can be used to track access to documents. It also tracks view accesses. Stellent is the product used for the IES Web FileShare (WFS) application. It makes use of the corporate metagroups for its access control and resides on both the SRN and the SCN.

Oracle

Oracle has a product called Oracle Files, which a 2004 Gartner report described as providing "light weight file management." A new product, Oracle Files 10g, will focus on enterprise content management and workflow capabilities. Both products are applicable to electronic data content. SNL currently does not own Oracle Files or Files 10g.

Serena TeamTrack

Serena TeamTrack provides a central point of access from a web browser for users who can submit, receive, or manage issues or requests. Users can view or report on issue status, anywhere, any time.

TeamTrack provides Issue and Defect Tracking (IDT) that involves collecting and managing issues and defects, starting at capture and continuing through to the resolution.

IDT, also known as bug tracking, is popular in software application development, hardware development, and more broadly in any process in which recurring issues need to be addressed.
When either internal testing or a customer report uncovers an issue or defect, the following process typically ensues:

- Issue logged
- Issue triaged
- Issue assigned
- Issue validated
- Issue resolved and packaged

Many entities may participate in this process: representatives from customers, end users, development, quality assurance, product management, marketing, technical support, and management. Resolution paths may vary depending on issue severity, type, impacted resources, applications, or customer.

Some of the Business and IT Benefits of using TeamTrack are:

- Customizable Work Flow that can be changed without work disruption
- Platform and device-neutral accessibility
- Real-time reporting and trend analysis
- Web-based interface
- Integration with Microsoft Outlook and Project
- Customizable templates
- Out-of-the-box integration
- Support of Web services

Serena TeamTrack is a corporately supported COTS software package that is widely used for many projects including several affiliated with YMP.

8.3 Potential Solution Recommendation

COTS analysis clearly illustrates that there is no single COTS product available that could replace TDMS and DIRS as a complete solution right out of the box. The customization effort required would significantly negate any savings in cost. However, there are COTS solutions (e.g., TeamTrack) that could be part of the solution

.

There seems to be a number of in-house solutions that are relatively compatible and could be partly reused. However, none of them is an out-of-the-box solution either, and will require several COTS software and hardware along with customization (i.e., Oracle database, eMatrix, ESAW server, and APC Client,).

DART SDM tools would not address requirements related to business rules, which would be handled by other software tools that interact with the DART SDM tools. Besides the non-functional requirement to control privileges based on roles and groups, other non-functional requirements are also not typically addressed by the DART SDM tools. Since the DART data management core is intentionally both generic and extendable, it can possibly satisfy many of the TDMS functional requirements, especially the ability to track input-to-output dependencies between artifacts, including the specific versions of artifacts involved in the relationship. However, the DART tool will still not be an out-of-the-box solution and will need to be customized and integrated with other tools and technologies.

An **AMR Exploration Tool** is being developed specifically to support TDMS reporting needs required for License Application defense.

SDDB has many similarities with TDMS. They are both organized around the three main functional areas (1) data characterization/entry, (2) data QA/QC, and (3) data retrieval/use. SDDB also has similarities in its reporting requirements of data traceability (see graphical representation in Appendix D). Because of SDDB's overall similarities to TDMS, much of its underlying environment could be reused. Most of all, there are skills and internal expertise available in implementing a TDMS-type application, and this knowledge and familiarity can be recycled.

Both the **Issues Management Database** and **Serena TeamTrack** have the capability to track issues that may be generated by the TDMS administrators and users. TeamTrack has the advantage of being a corporate supported tool with a dedicated team of administrators. In addition, TeamTrack is a COTS package with support provided by the vendor. The Issues Management Database has the advantage of already being deployed and in use by some of the TDMS users and would require little modification to be incorporated into the TDMS redevelopment effort. The distinct disadvantage is that it is an Open Source software package with no support from a vendor. Additionally, the current deployment is not within a corporately supported infrastructure. Further analysis will be required in the design phase to

determine which product can be supported within the Yucca Mountain corporate supported infrastructure.

The best strategy we can recommend would be a combination of in-house development, while at the same time leveraging the re-use of Sandia tools and COTS. Advantages include:

- Delivering custom applications built to fit to YMP business processes and goals instead of modifying YMP business processes around the packaged software solution
- Existing tools can be re-used lowering overall development costs while improving quality
- Instead of one big drop after development, there is the possibility of smaller, iterative drops for user testing and feedback with the in-house staff providing tighter control of the project
- Levering multi-application data integration can require significant ongoing administrative effort, which can be reduced with custom tools or reports
- Control over delivery schedule especially given the short timetable (next license application is scheduled for 2008)

Formatted: Bulleted + Level: 1 + Aligned at: 0.25" + Tab after: 0.5" + Indent at: 0.5", Tabs: 0", List tab + Not at 0.63"

9.0 DATA MIGRATION ISSUES AND RECOMMENDATIONS

Issue: Inconsistent data in Ingres databases

Data entered according to business rules are no longer in effect:

- DIRS – concept of TBDs is now defunct, but there are still references in the database coded as TBD
- TDMS – SEP data was stored in the SPA databases until the decision was made to store it externally to databases and store only the links to the data
- DIRS – Document status in working documents does not reflect status of document in CDIS

Formatted: Indent: Left: 0.25", Bulleted + Level: 2 + Aligned at: 0.75" + Tab after: 1" + Indent at: 1", Tabs: Not at 1"

Our recommendation is that an analysis report be provided to the data owner (delegated by YMP management), who will determine which of the following options to pursue:

- Store the data in the original format in a separate database, which is set up to hold "legacy" data
- Unload it into ASCII format and store the files in RIS with a brief description of the data
- Modify data to a new format

Formatted: Indent: Left: 0.25", Bulleted + Level: 2 + Aligned at: 0.75" + Tab after: 1" + Indent at: 1", Tabs: Not at 1"

Issue: Orphaned data files

These are files that do not have a valid URL pointer or pointers with no valid file at the location:

- URL may not reflect the true location of a data file.
- A file may not have a URL link

Formatted: Indent: Left: 0.25", Bulleted + Level: 2 + Aligned at: 0.75" + Tab after: 1" + Indent at: 1", Tabs: Not at 1"

- A file may have the same link as another file

We recommend that orphaned files, that is, files without a valid URL pointer (link) be stored in RIS with a brief description of the file and its location. A data analyst (delegated by YMP management) should be responsible for this.

Orphaned Pointers, that is, URL links without an available target file need to be reported to a data analyst (delegated by YMP management), who will trace the ownership back to where the data was created (either internal or external entities). A replacement should be sought where possible. In the event this is not possible, a YMP data analyst should trace the data usage to determine the impact and recommend re-working or removing the URL from references

Duplicate URL links, that is, files with more than one URL link need to be reported to the data owner

Audits should be run regularly to identify orphans and duplicates. The results should be reported to a designated data analyst who would determine the disposition of the data (as noted above).

Issue: Old data in databases are no longer referenced

Work on them has been suspended or canceled upon changes in DIRS System.

- TAG Records were implemented in the database but there is no supporting application to manage and report them.
- Some status values were available in the past, but are no longer valid.

Formatted: Indent: Left: 0.25", Bulleted + Level: 2 + Aligned at: 0.75" + Tab after: 1" + Indent at: 1", Tabs: Not at 1"

We recommend not converting TAG records – TAG records were to support administrators in their work and were never implemented.

A report on invalid values should be generated for a designated analyst who will determine the possible disposition – (1) decode to new status, (2) do not convert (omit), and (3) extract data and save it in RIS.

Issue: Some legacy data elements in the DIRS and ATDT databases will not have a direct mapping to new data stores:

A new database design will not contain the same attributes as the current database design, and therefore some elements of the current databases will not have a corresponding element in the new design.

Our recommendation is to:

- Analyze existing data sources and map the current format to the appropriate target location.
- Identify elements that can not be mapped.
- - Determine and document their disposition, that is, (1) do not convert, or (2) capture elsewhere and store as legacy data
- Utilize mapping matrices to document mappings for new values

Issue: Some data are in files (various formats) that are not available in current applications and have no defined disposition.

- Legacy data may or may not have future relevance.
- Business owners are reluctant to delete this data.
- Format may be from databases that are old (unsupported Ingres).

Formatted: Bulleted + Level: 1 + Aligned at: 0.25" + Tab after: 0.5" + Indent at: 0.5", Tabs: Not at 1.5"

Our recommendation is to generate a report of the data files and assign responsibility to YMP management to analyze the relevancy of each file:

- For relevant data, store files in RIS with a brief description.
- For non-relevant data, obtain data owner's approval to delete data

Issue: Risk that not all data will be converted

Our recommendation is to employ sound conversion practices that include:

- Analyze data to identify data that can not be converted (bad data values, etc.)
- Obtain corrections from data owners where possible
- Employ data cleansing procedures to correct data
- Develop migration scripts that can be repeated
- Employ migration tools to assist in conversion (Example: SQL Server SSIS or DTS, Data Stage – data warehouse tool)
- Develop audit process to account for all data converted or not converted
- Document migration process
- Test migration process
- Review audit results and migration process with data owner

Formatted: Bulleted + Level: 1 + Aligned at: 0.25" + Tab after: 0.5" + Indent at: 0.5", Tabs: Not at 1.25"

9.1 Infrastructure Constraints

Our assumption is that new DIRS and TDMS will be deployed on the new Sandia YMP network and managed by Sandia IT.

Issue: Infrastructure Constraints

Applications deployed and managed by Sandia IT must follow Sandia IT standards for application development methodology, development tools, database management, and operations.

- Follow software engineering practices, that is, utilize development, quality, and production environments.
- SILC is the applications development methodology
- J2EE is the development standard for creating application code (current products supported are WebLogic and ColdFusion)
- SQL Server and Oracle are the only database management platforms supported
- Operations standards include:
  - Control M/ECS for scheduling work
  - Help Desk for managing outages and reporting problems
  - Network operations for monitoring scheduled work
  - Service level agreements for application availability and support services (Systems Administration, Network, Database Administration)

Infrastructure required to support development, deployment, and operation of TDMS/DIRS applications:

- Software development environment (Web Logic and Cold Fusion servers)
- Database Management system environment (SQL Server 2005 on Windows 2003 or Oracle 10g on HP Unix data base servers)
- Operations environment (ControlM scheduling software, Network Operations monitoring, etc.)
- Development, quality, and production environments for both application and database servers

The current Sandia YMP environment does not provide the required infrastructure:

- Current support is for basic services (e-mail, SharePoint, desktop office products)
- No plan to provide complete infrastructure for development and deployment (as defined above)

Since deployment of new DIRS and TDMS on the Sandia network will still require access to data and applications on the Bechtel network, managed by Bechtel IT, deployment and development must be on the Sandia YMP network.

TDMS/DIRS users expect that performance will be very high, including sub-second response times.

<u>Our recommendations for a path forward include:</u>

- Add missing infrastructure components to Sandia YMP network

- Web Logic/Cold Fusion application servers
- Database Servers (SQL Server or Oracle)
- ControlM/ECS scheduling agent
- Add operations monitoring of YMP components to existing Sandia center in Albuquerque
- Contract with Sandia organizations to provide support services for database administration, systems administration, and help desk services for TDMS/DIRS applications
- Provide access to Sandia YMP network from Sandia SRN or SON Networks.
- NOTE: Database and application servers must be on YMP networks because response times cannot be guaranteed when data must cross networks (from Sandia SRN to Sandia YMP).

- Provide access to Bechtel Network from Sandia YMP Network

  - Negotiate access to Bechtel databases
  - Copy required Bechtel data to Sandia YMP Network.

- Contract with Sandia Lab to develop and implement new DIRS and TDMS application:

  - Sandia development organization is skilled at developing applications that meet required Sandia IT standards
  - Sandia Database Administration organization is skilled at supporting Sandia developed applications
  - Sandia Database and Systems Administration can provide necessary service levels – currently provide 24X7 support to existing applications

- Develop new TDMS/DIRS applications on Sandia YMP Network:

  - TDMS/DIRS still requires access to Bechtel managed data
  - Development in Sandia Albuquerque network does not test performance response times in Sandia YMP

Issue: Non-Infrastructure Constraints

- The next phases of the TDMS/DIRS development project will require involvement of Bechtel IT Staff:
  - Data conversion of existing Ingres data bases
  - Data migration from applications that still need to interface with TDMS/DIRS
  - Access to applications that still need to interface with TDMS/DIRS
  - Access to software drivers to access old unsupported Ingres data bases

Formatted: Bulleted + Level: 1 + Aligned at: 0.5" + Tab after: 0.75" + Indent at: 0.75", Tabs: Not at 1"

- The next phase of the TDMS/DIRS development project will require more involvement with the TDMS/DIRS user and administration community.

- Prepare test cases to validate new system
- Verify conversion effort, identify and resolve conversion data problems, validate and approve application design
- Validate system functionality

Formatted: Bulleted + Level: 1 + Aligned at: 0.5" + Tab after: 0.75" + Indent at: 0.75", Tabs: Not at 1"

We recommend the following:

- Negotiate with Bechtel IT Staff support during next phases of development:
  - Provide funding support for conversion, migration, testing, and deployment
  - Negotiate service level agreement

Formatted: Bulleted + Level: 1 + Aligned at: 0.5" + Tab after: 0.75" + Indent at: 0.75", Tabs: Not at 1.25"

- Identify key individuals necessary for project and add them to development team
  - Identify key individuals from user community
  - Identify key individuals from TDMS/DIRS administration community
  - Identify key management responsible for TDMS/DIRS
  - Add them to project team, and make managers responsible for the success of the project
  - Cross train backup teams to take over some work load during peak project periods

Formatted: Bulleted + Level: 1 + Aligned at: 0.5" + Tab after: 0.75" + Indent at: 0.75", Tabs: Not at 1.25"

10.0 TERMS

Term	Description
Abstraction	Distillation of the essential components of a process model into a suitable form for use in a total system performance assessment: The distillation must retain the basic intrinsic form of the process model but does not usually require its original complexity. Model abstraction is usually necessary to maximize the use of limited computational resources while allowing a sufficient range of sensitivity and uncertainty analyses
Acquired Data	Data that are obtained as a result of data-gathering activity: Acquired data may be procured or obtained form outside project sources or recorded as a result of a project field or laboratory data-gathering activiy. Project field or laboratory data recorded as raw data but converted to scientific or engineering terms are acquired data.
AFS	Automated Form System: Provides and manages the current approved versions of forms
AMR	Analysis Model Report: Documents that describe the individual analysis models and how the respective parts of the repository work: Contains the more detailed technical information used to support TSPA. Consists of data, analyses, models, software, and supporting documentation that will be used to defend the applicability of each process model for evaluating the post-closure performance of the potential Yucca Mountain repository system. These models simulate the different geologic, hydrologic, physical, and chemical processes of the repository.
ATDT	
BSC	Bechtel SAIC Company: The previous YMP Lead Laboratory
CDIS	Control Document Information System: Stores control documents, i.e., AMR, calculations, desktop information and change documents
Data or technical data	Information measured or derived from scientific investigation activities both in the field and the laboratory: Parameters that have been derived from raw data are sometimes considered to be data.
Data Coordinator	Individual trained to submit data to TDMS after the data are compiled and formatted by the data originator or data preparer
Developed Data	The results of reducing, analyzing, or interpreting data after data acquisition: Developed data are composed from source data taken from controlled sources such as data in the TDMS, document(s) retained in the Technical Information Center (TIC) and scientific analyses and models
Direct Input	Those inputs used in a technical product that are directly relied upon to support the results or conclusions
DIRS	Document Input Reference System: The database used for recording and tracking technical product input (direct and indirect)
DIRS Report	A compilation (usually in paper form) of all cited technical product inputs related to a technical product, which is generated from the relevent DIRS Working Document.
DIRS Working Document	An electronic document created in DIRS in order to catalog, track, and update cited technical product inputs for a technical product.
DOE	Department of Energy
DTM	
DTN	Data Tracking Number: A unique identifier assigned by the Data Coordinator to each data item tracked in the Automated Technical Data Tracking Database
DTNRS	Data Tracking Number Request Sheet
FEP	features, events, and processes
Indirect Input	Inputs of a technical product that only provide supporting information, and are not used in the development of results or conclusions in the technical product (e.g., model validation, sensitivity studies, and neutralizations)
IRAN	Impact Review Action Notification:
LA	license application
LSN	
NSHE	Nevada System of Higher Education
NRC	Nuclear Regulatory Commission
OCRWM	Office of Civilian Radioactive Waste Management
Originator	The individual who signs a technical product as Originator/Preparer: With respect to this procedure, the functions assigned to the originator may be delegated to a designated editor.
OSIT	Office of Science and Technology

Term	Description
PITT	Parameter Tracking Tool: Incomplete and currently not being used
Q System	Quality Affecting System
QA	Quality Assurance:
QAP	Quality Assurance Program
QAPP	Quality Assurance Plan Procedure
QARD	Quality Assurance Requirements Document
Qualified Data	Data collected under an approved QA program that meets the requiremetns of 10 CFR 63.142 (or previously implemented 10 CFR 60 QA program) (i.e., qualified from origin), or unqualified data that have undergone the qualification process
RIS	Record Information System
RTM	Serena® Requirements Traceability and Management
SCM	Software Configuration Management Tool
SEP	Site and Engineering Properties Database A database component of GENESIS, which contains site investigation field and laboratory test data, as well as engineering analysis input data, organized by parameter, or Specific Data Tracking Number. technical data obtained through the data collection activities in support of the various facets, documents, and so forth, for the Yucca Mountain Project.
SNB	Scientific Note Book
SPA	
SUBCAT	Submittals Catalog Application: SUBCAT is designed to work in conjunction with the SEP database. The SEP database is intended to be a repository. The SUBCAT application is a cataloging system with its primary purposes being twofold: (1). an indexing system (keeping track of the data submittals and allowing retrieval of such data submittals upon demand); and (2). an input tool of data for incorporation into SEP (a "data pre-processor used in the creation of tables to contain the data, table descriptions, footnotes, specific information of each of the data columns, information regarding the parameters and attributes assigned to each of the data columns, and so forth).
TBV	To Be Verified: A unique number assigned to a reference (DTN, under -development information, vendor data, or assumptions, etc.) that requires approval before the reference can be qualified for use
TDIF	
Technical Product	Any product developed by or for the YMP containing information with relevance to the characterization
TIC	(Technical Information Center)
TSPA	(Total System Performance Assessment)
Unqualified Data	Data developed prior to the implementation of 10CFR 60,Subpart G, QA Program, or data developed outside the Yucca Mountain Project such as the oil companies, national laboratories or universities, or data published in technical or scientific publications. It does not include established fact.
USGS	(U.S. Geological Survey)
YMP	(Yucca Mountain Project):

11.0 REFERENCES

1. U.S. Department of Energy, Revised Interim Guidance Pending Issuance of New U. S. Nuclear Regulatory Commission Regulations (Revision 01, July 22, 1999), for Yucca Mountain, Nevada (Dyer, 1999) Environmental Protection Agency amended standard??
2. DOE/RW-033P, "Quality Assurance Requirements Description" (QARD), and
3. Attachment 1 "Quality Assurance Requirements for Work Authorized by OCRWM Program and Funding Guidance Memorandum".
4. Sandia Quality Assurance Program Plan (QAPP), *Supplement V, Control of Electronic Management of Data.*
5. SNL: IM-PRO-002, Control of Electronic Management Information
6. SNL: IM-PRO-003, Software Management
7. SNL: IM-PRO-005, Software Independent Verification and Validation
8. SNL: IM-PRO-006, Independent Verification and Validation
9. SNL: SCI-PRO-002, Records Management
10. SNL: SCI-PRO-004, Managing Technical Product Inputs
11. SNL: TST-PRO-003, Scientific Notebooks
12. YMP: TST-PRO-001; Submittal and Incorporation of data to the Technical Data Management System
13. YMP: LP-SIII.2Q-BSC; Qualification of Unqualified Data Revision 0 ICN
14. CPSR001.3: Integrated Laboratory Management System (ILMS)
15. CPR001.3.2 Quality Assurance
16. CPR001.3.6 Corporate Engineering Excellence
17. DOE Order 414.1C: Quality Assurance
18. DOE/NNSA Weapons Quality Policy (OC-1)
19. Software Lifecycle (SILC)
20. Total System Performance Assessment (TSPA)

APPENDIX A: DOCUMENT GRAPHICS

Only a few of the graphics referred to in the body of this document are actually presented there; the majority of them are presented in this appendix. This is due primarily to the complexity of the relationships within many of the graphics that may require time and study to digest. Rather than disrupt the overall flow of the analysis by distracting the reader with the intricacies of the graphics, the figures are collected here for later study. All graphics are provided here; both those contained and those simply referenced in the body of the text.

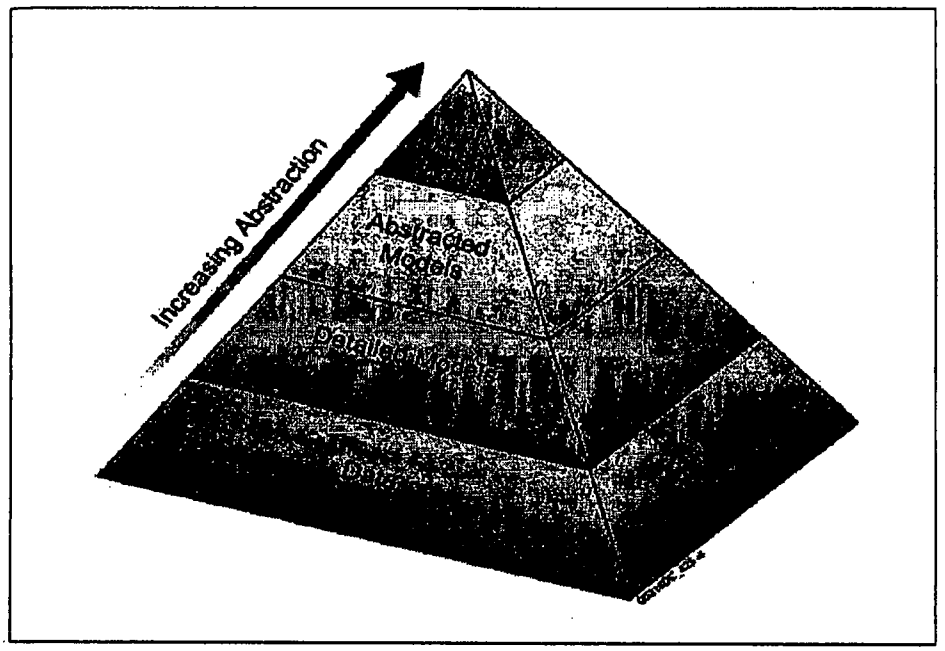


FIGURE A1. *Steps In Developing a Total System Model*

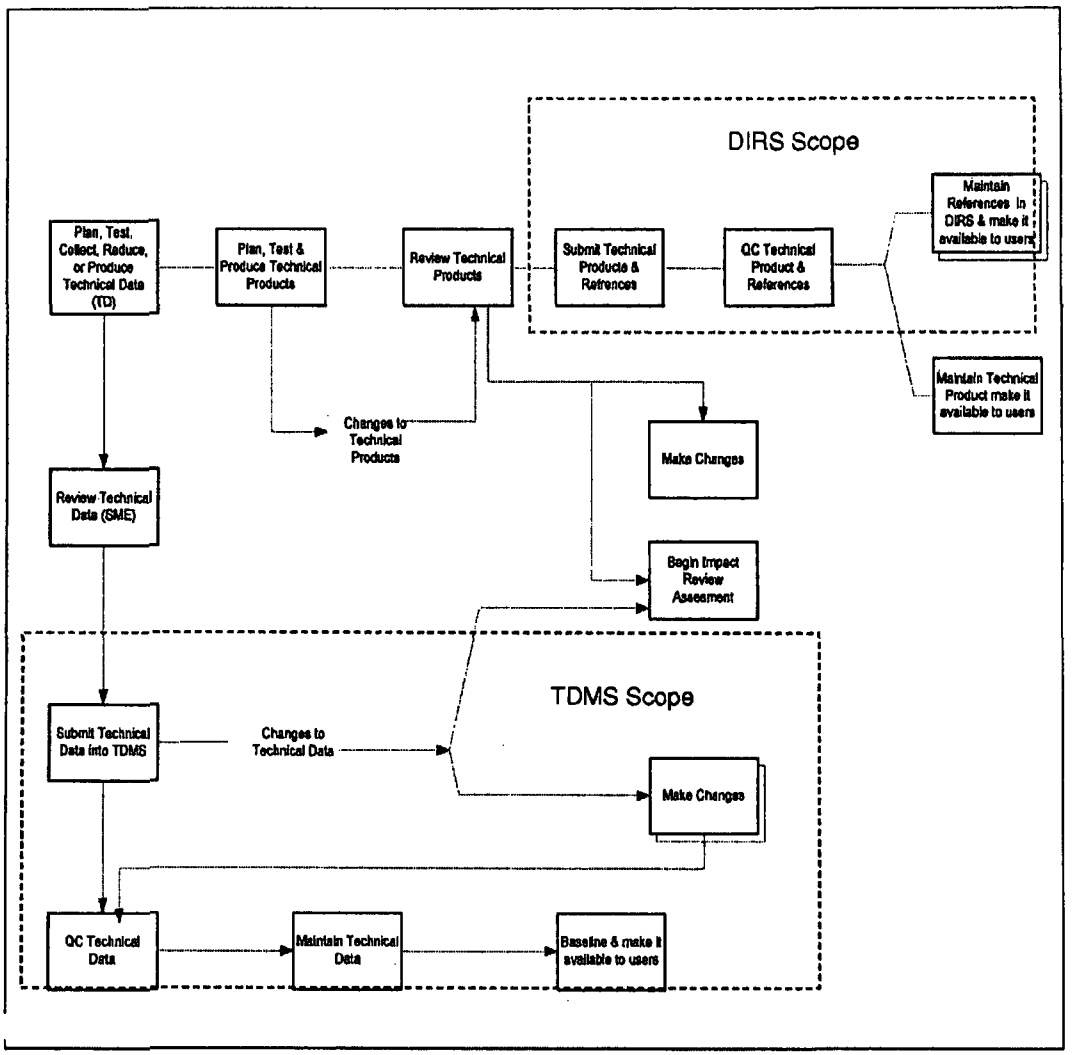


Figure A2 *Overview of Business Function: Scientific Investigation Process*

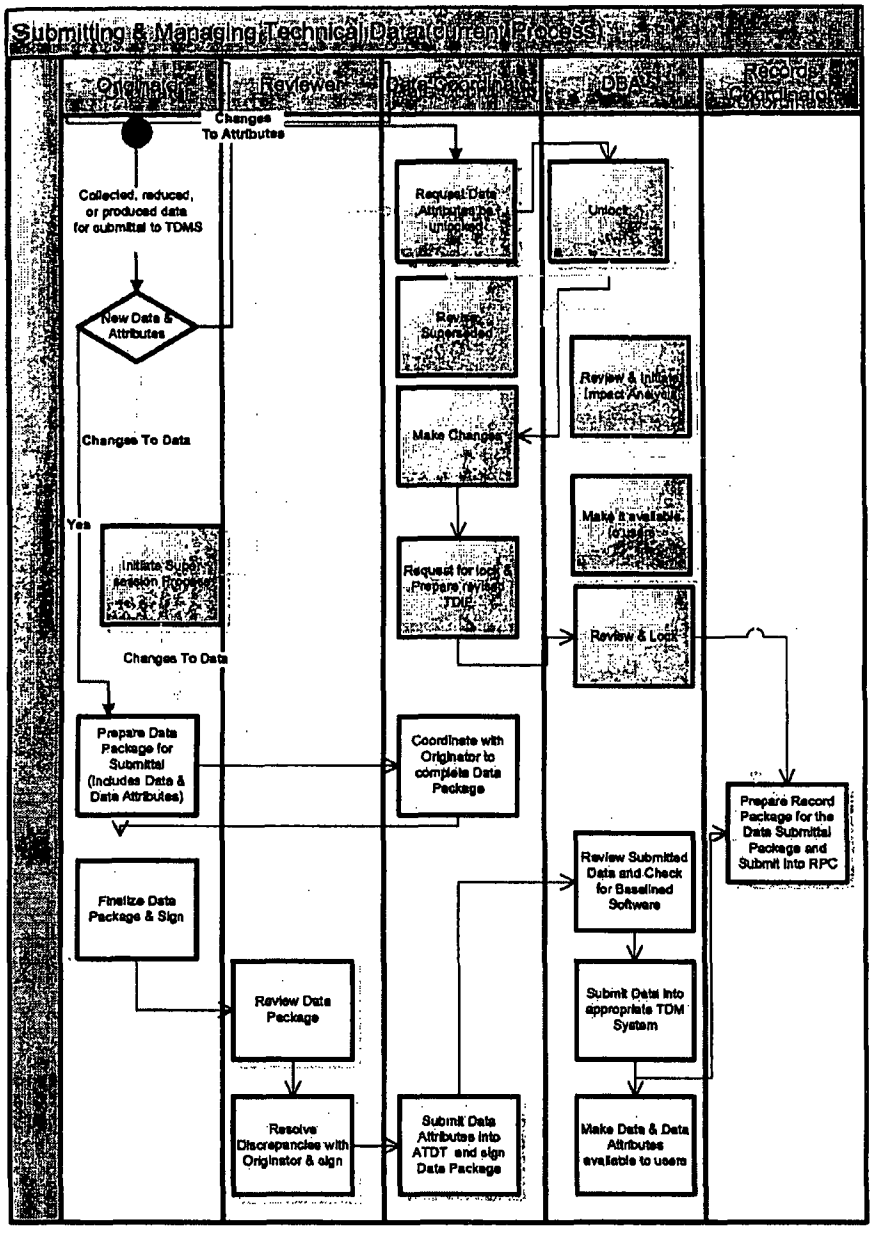


Figure A1 *TDMS Data Submittal and Management Process ("As-Is-State")*

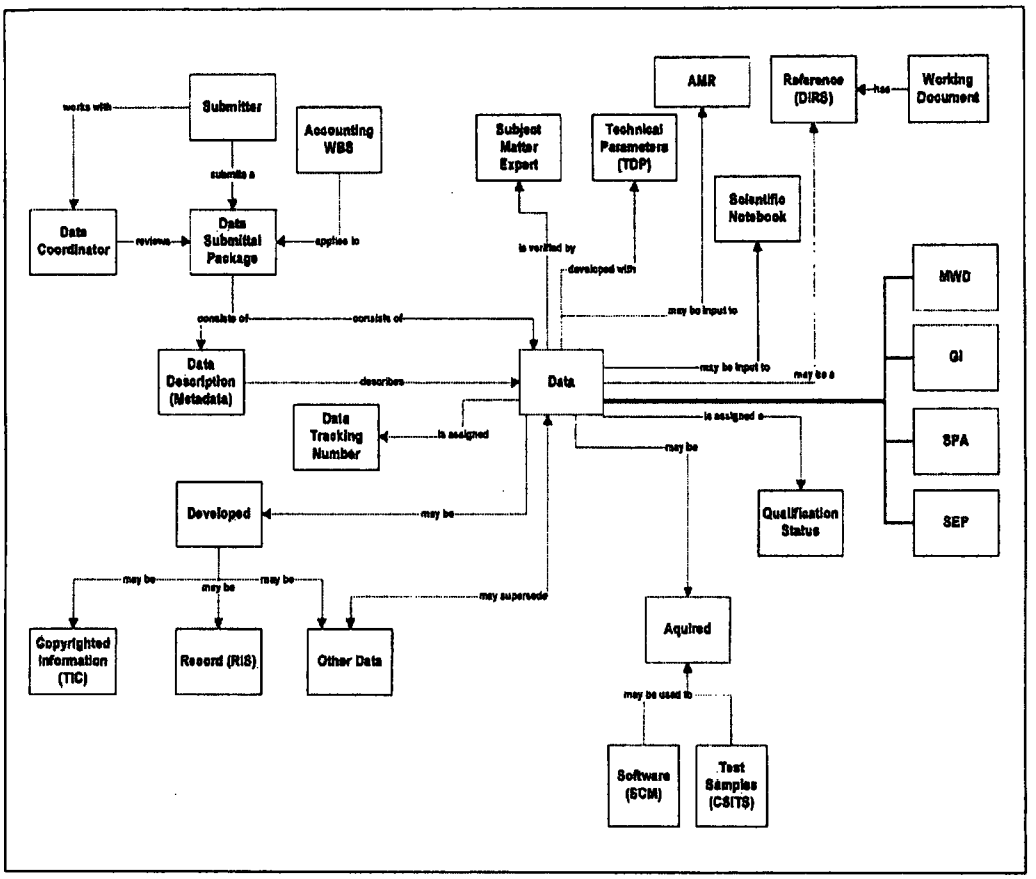


Figure 18. TDMS Fact Model ("As-Is-State")

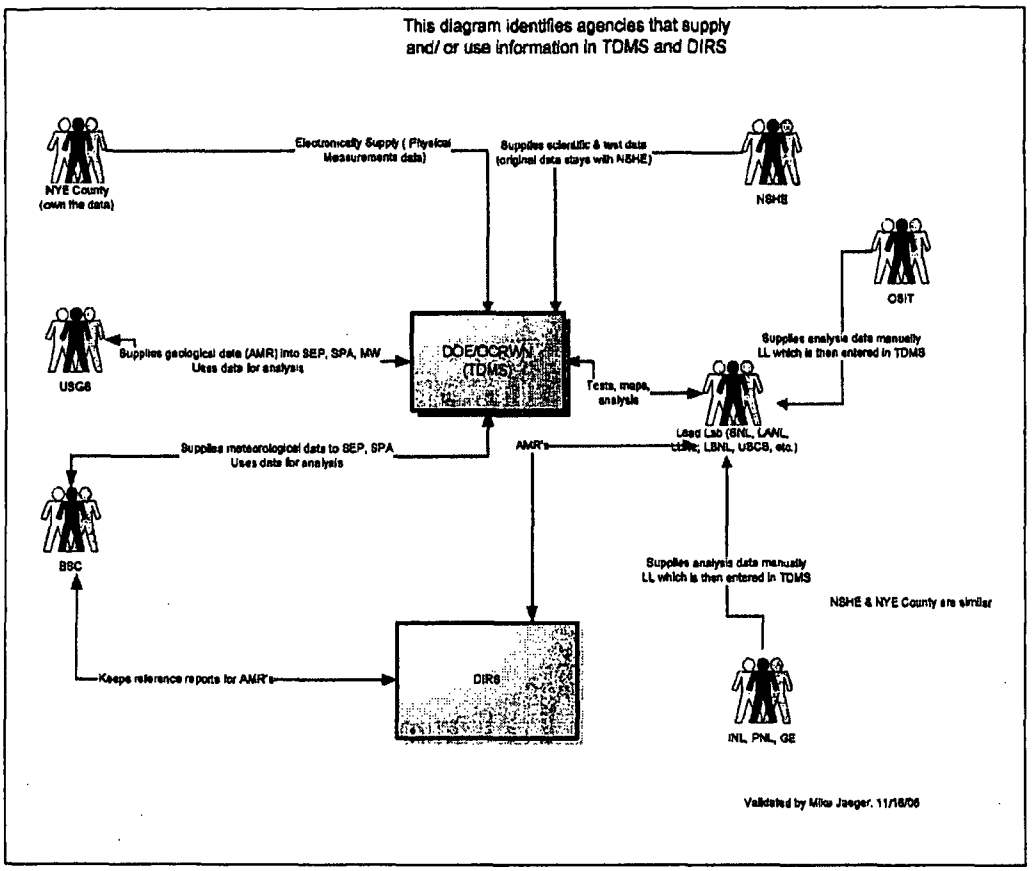


Figure A3. *OCRWM-Wide Organizations/Agencies Using TDM Systems*

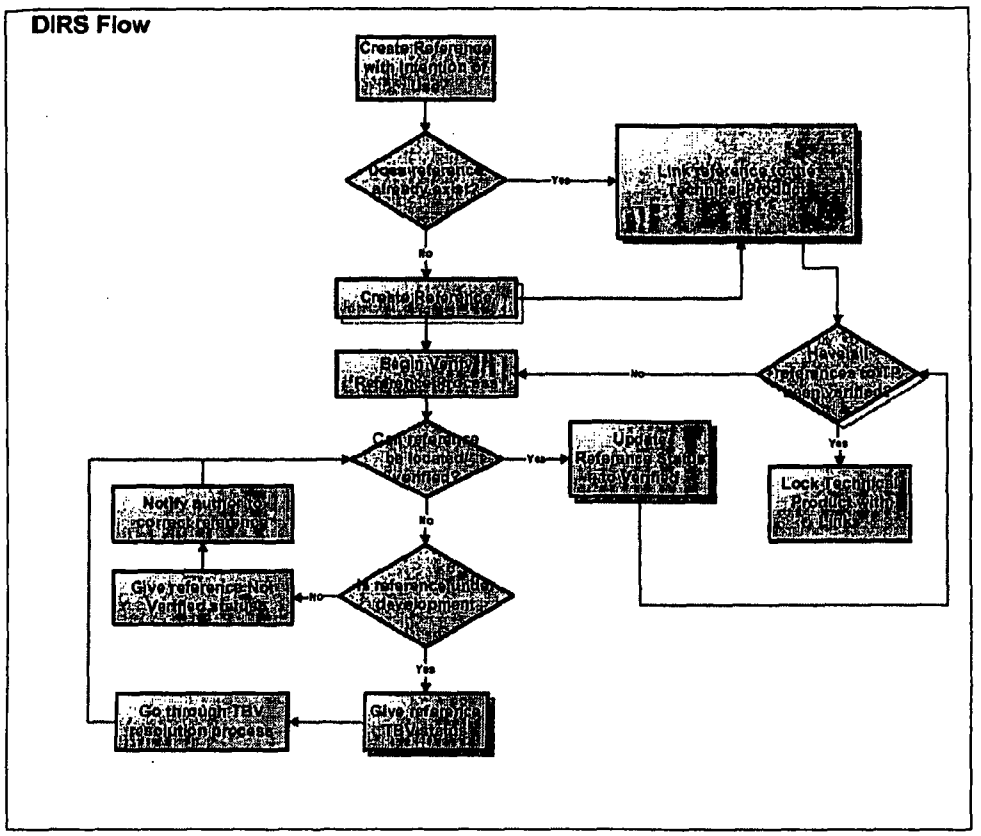


Figure A61. *DIRS Flow Diagram ("As-Is-State")*

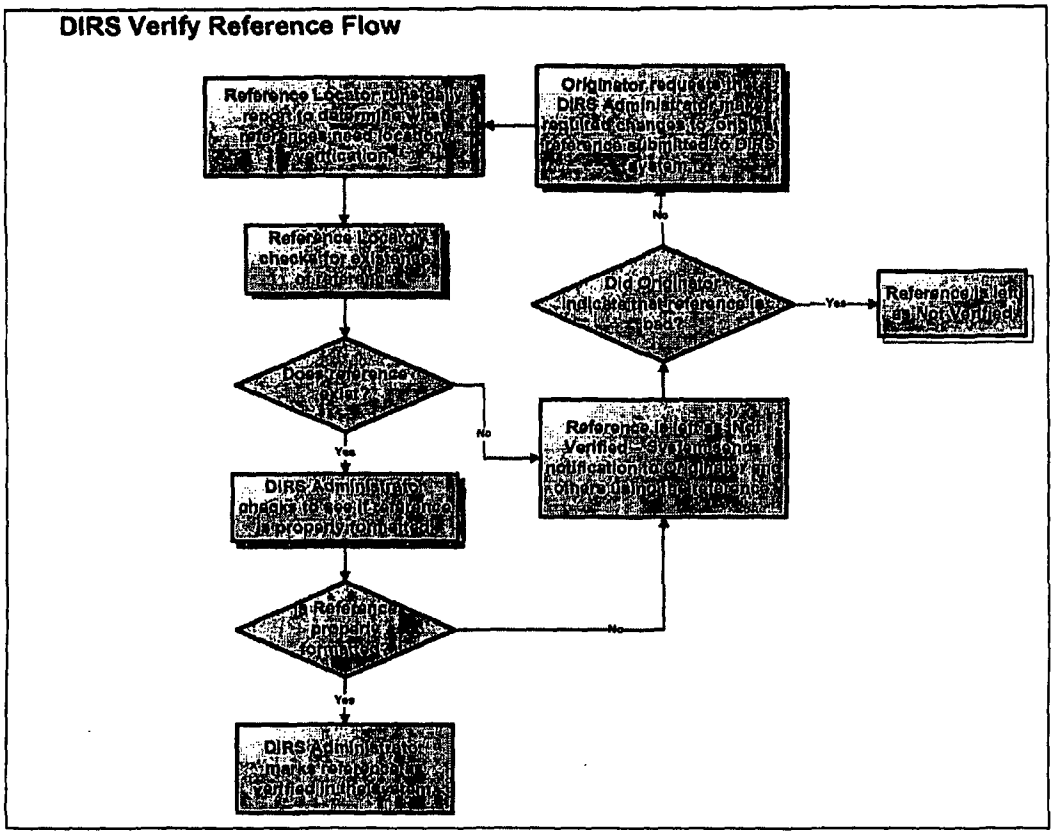


FIGURE A6. DIRS Verify Reference Process ("As-Is-State")

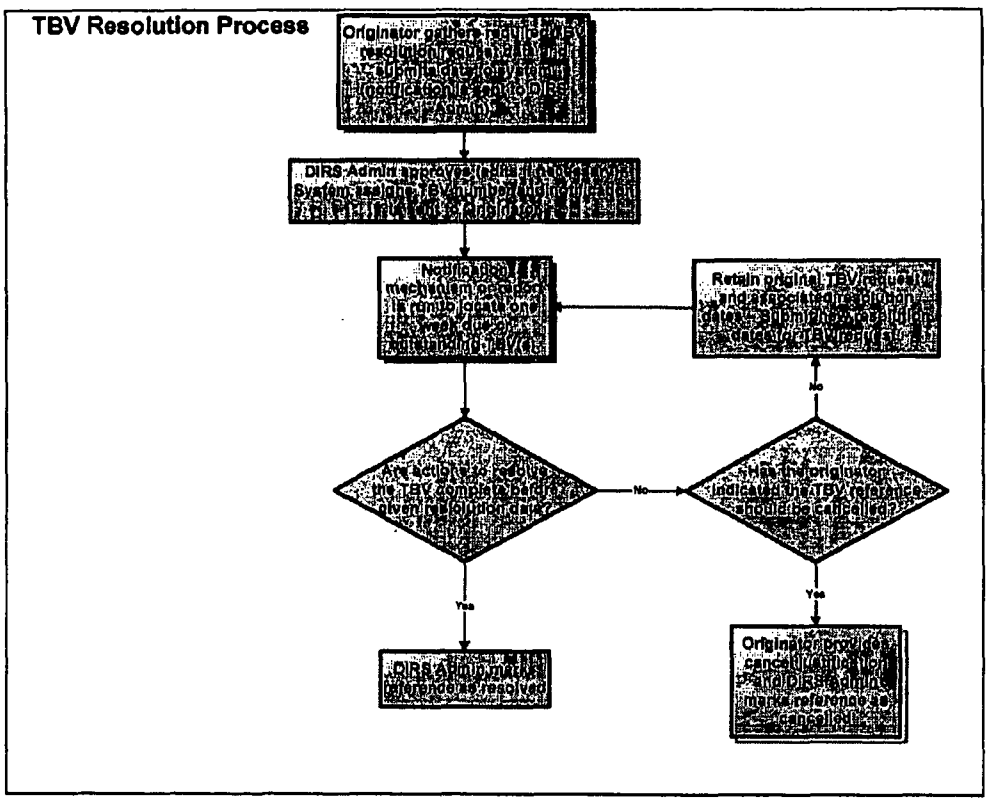


Figure A8a. DIRS TBV Resolution Process ("As-Is-State")

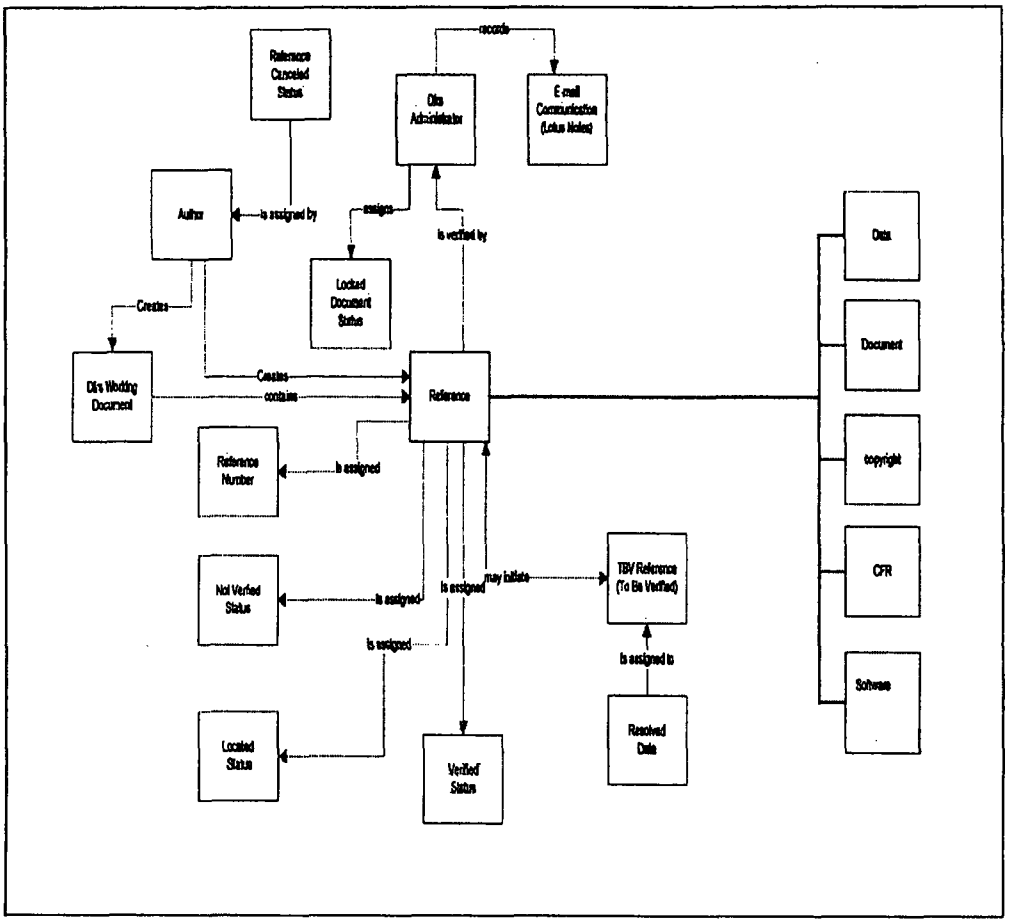


Figure 3. DIRS Fact Model ("As-Is-State")

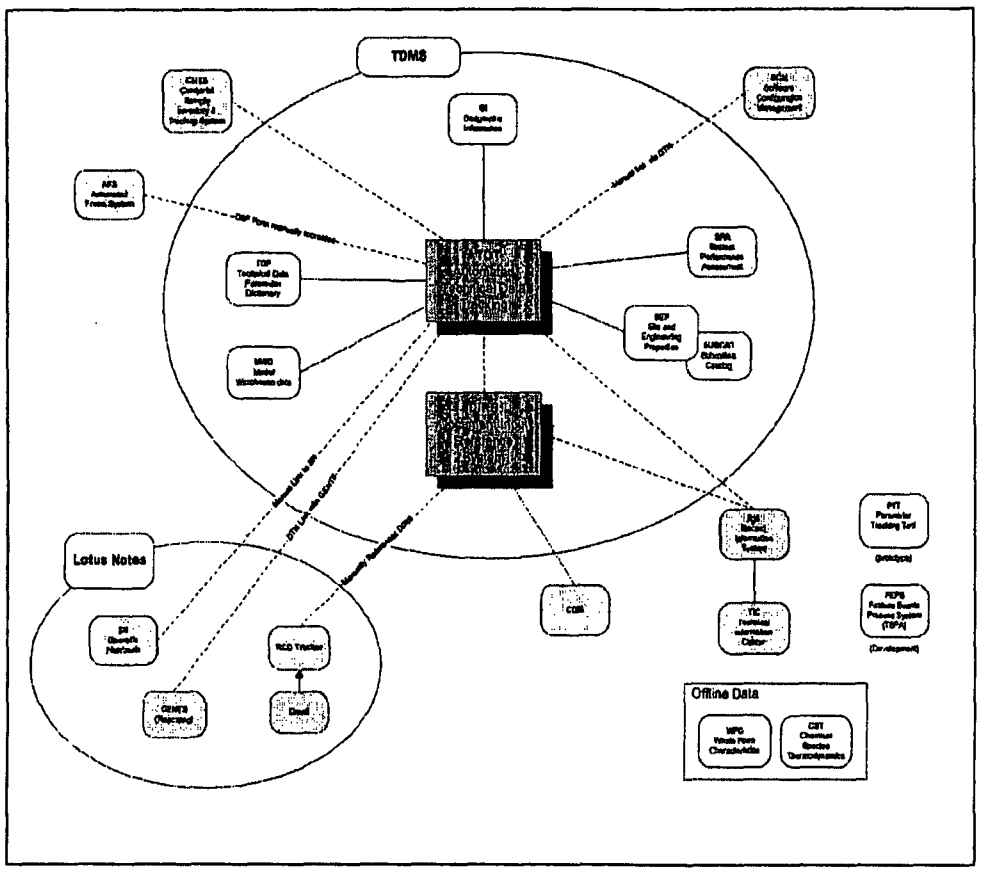


Figure A8. Overview of Technology: IT Systems Diagram ("As-Is-State")

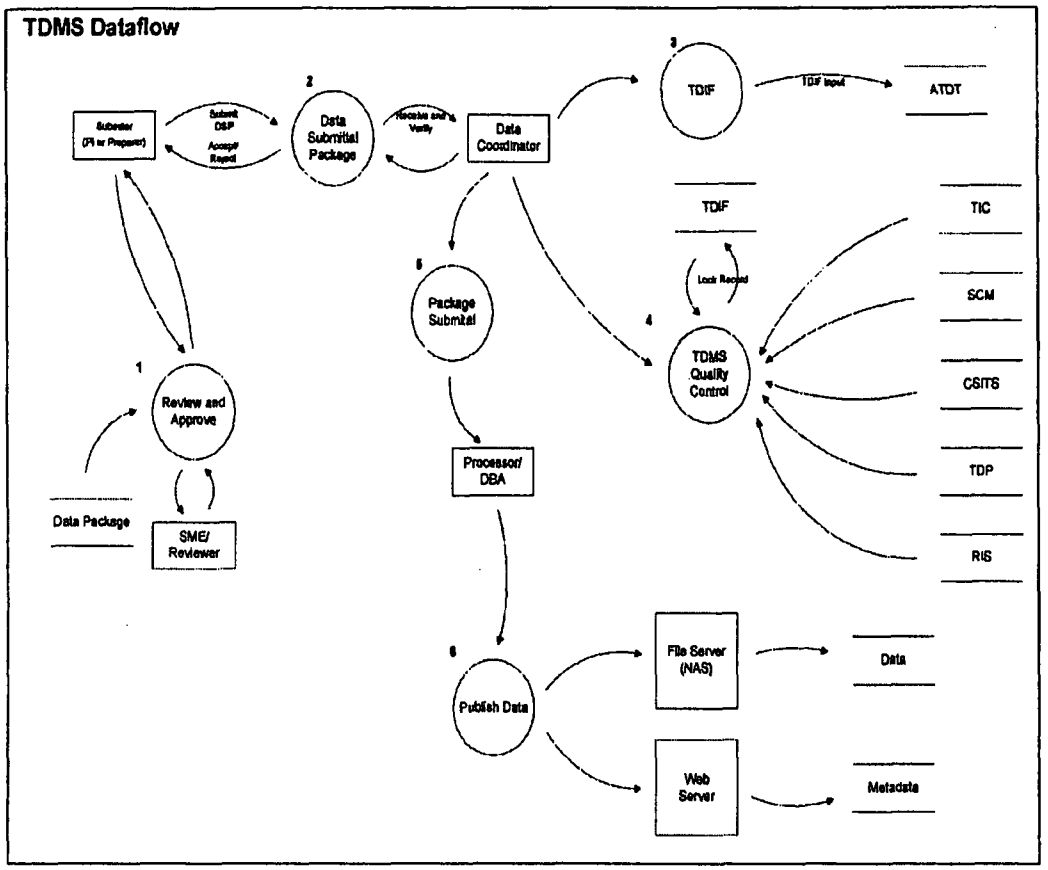
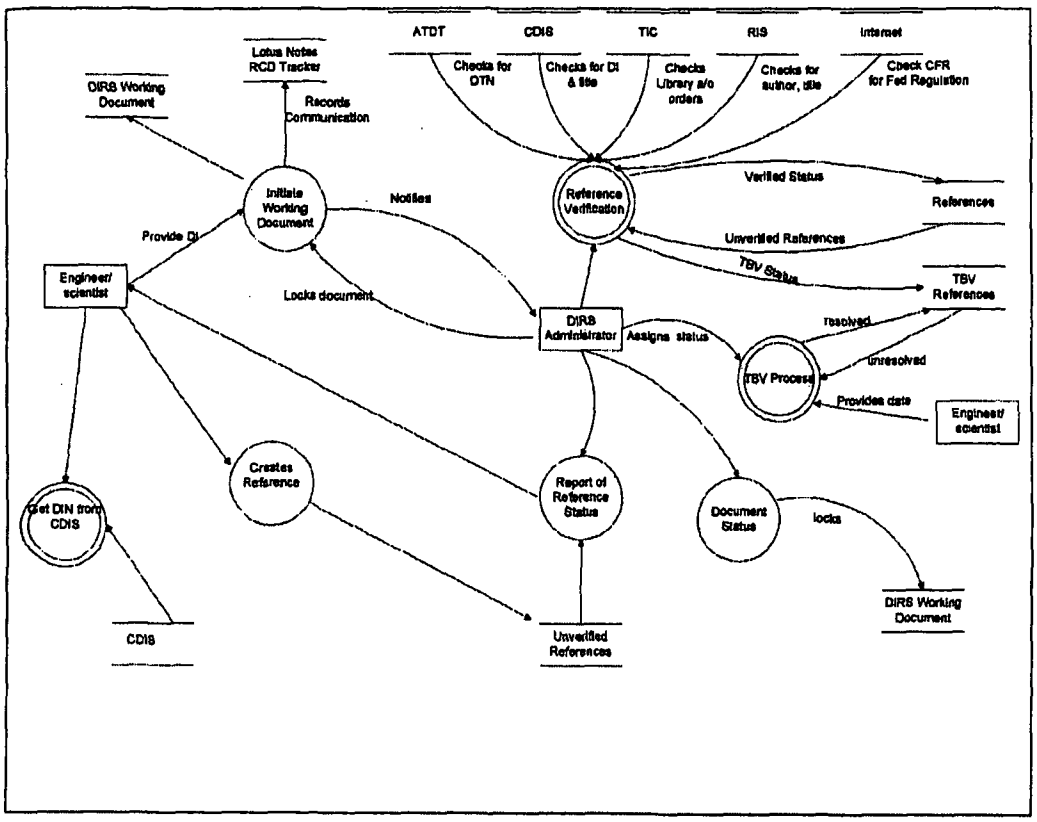


Figure A8. TDMS Dataflow Diagram ("As-Is-State")



[redacted]. *DIRS Dataflow Diagram ("As-Is-State")*

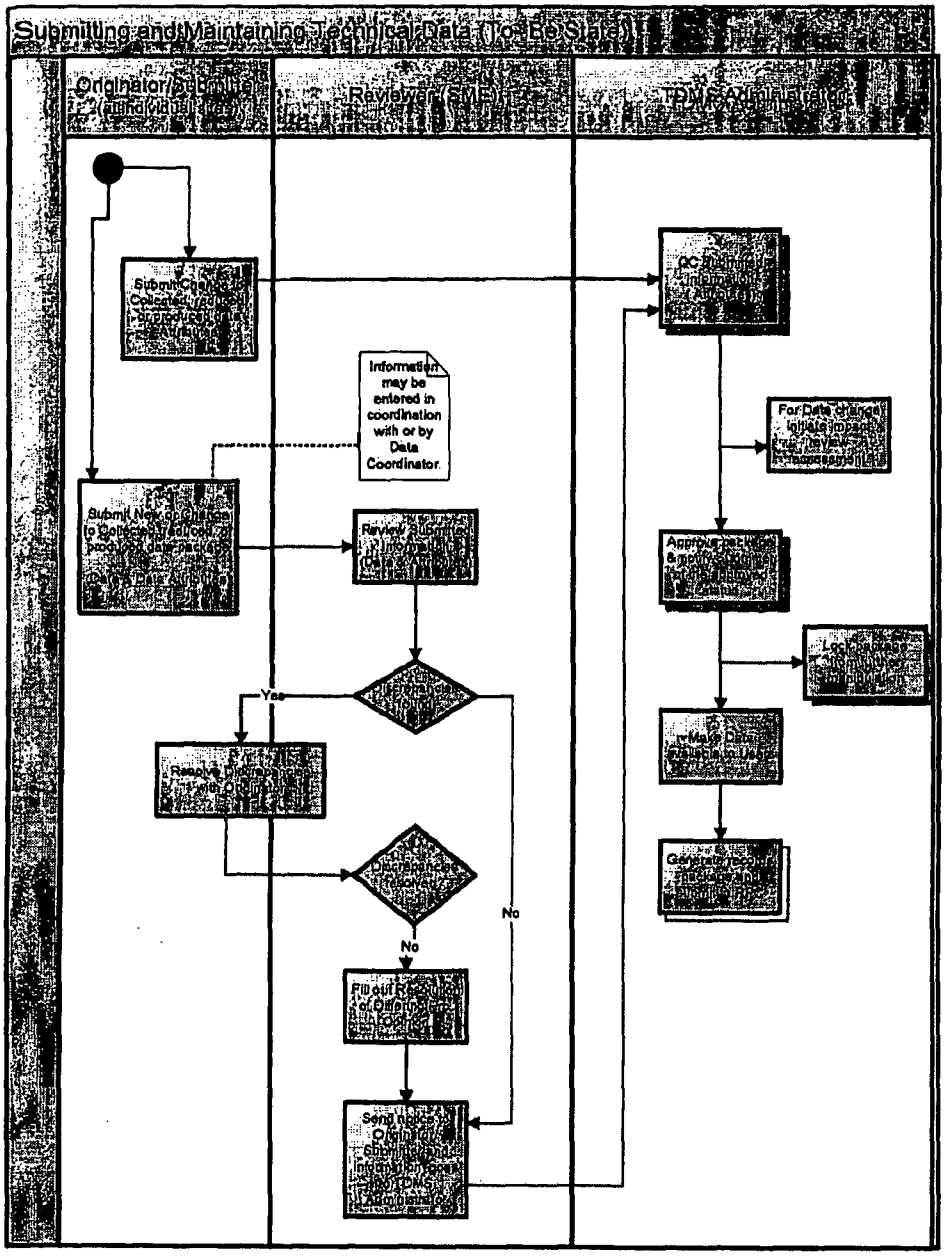


FIGURE A-1. Future TDMS Process ("To-Be-State")

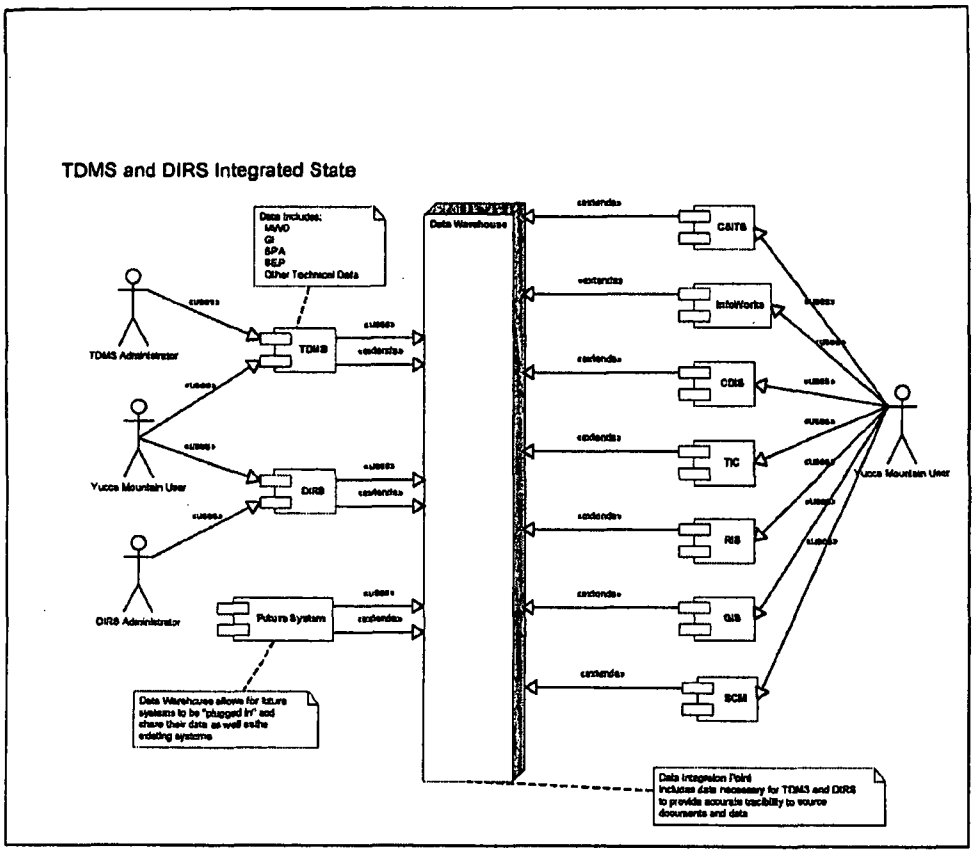


Figure A-12. TDMS and DIRS Integrated State ("To-Be-State")

APPENDIX B: POTENTIAL DATA WAREHOUSE CONTRIBUTION TO TDMS

Executive Summary

A data warehouse (DW) is an enterprise-level information repository dedicated exclusively to facilitating the analysis of business information and decision support, and a key technology for best business intelligence practices. Typically, a DW draws from various databases and operational systems throughout a company to provide a central repository where analytical queries and meaningful, integrated reporting can be performed.

There are many benefits to building and owning a DW; a few of which are listed below:

- No impact to operational system by queries to the DW
- Check operational system accuracy
- Analyze and execute business decisions based on data from multiple sources
- Submit queries spanning several operational systems
- Less time searching for and more time acting on information
- Exchange of information between different enterprises
- Summarize, view, and report the same set of data in different ways at different levels by different groups of users

These benefits can be used and applied in addressing some of the issues and gaps that have been identified in the Yucca Mountain Project (YMP) Technical Data Management System (TDMS). Various ways that a DW can be beneficial in supporting the licensing application (LA) for the Yucca Mountain repository are summarized below with more detailed explanations provided in later sections of the document.

- **Integrating disparate data into a single repository.** The data that currently reside in various disparate systems outside of TDMS (such as Lotus Notes, CDIS, RIS, TIC, SCM, and CSITS) can be integrated into a central repository. Combined, the data are easily accessible to TDMS subcomponents (such as ATDT and DIRS) to capture, track, and maintain traceability for references in major YMP documents.
- **Reducing manual processes.** With a DW in place, the manual processes that are being used currently to track pieces of information both inside and outside of TDMS specific systems will be reduced substantially because information can be obtained from a central data repository.
- **Enhancing data quality and integrity.** The integrity and quality of YMP essential scientific and engineering data would be enhanced since part of the warehouse building process includes checks to detect inconsistent and duplicate data. This requires that the data be cleaned up in the source systems before being loaded into a DW.
- **Enabling queries that cut across different databases.** The need to have stronger mechanisms within TDMS to pull data from various data sources can be addressed via the DW. Through a central data repository, queries and custom reports could easily be built and maintained, thereby decreasing the workload on operational systems.

- **Providing a platform for analytical queries and integrated reporting.** A DW repository integrates the data from all of the associated YMP systems to provide users with a platform upon which querying, reporting, and the use of analytical tools can be based. For instance, the ability to obtain and report on the qualification status of a DTN can be easily attained through the warehouse.

Information is a valuable asset. A properly designed and implemented DW can be a valuable tool for managing and using that asset. It brings together the vast volumes of detailed, unorganized data that are captured via the operational system and transforms them into a more unified consortium of data that can provide useful feedback, predictors, and warnings that help users at every organizational level make more informed decisions.

1.0 INTRODUCTION

The implementation of a DW at the YMP would enhance management's ability to analyze existing data relevant to the LA process. An effort to redesign the current TDMS is underway; however, the scope of that effort does not include all of the external systems needed to support an end-to-end solution. Several key elements will still be tracked and maintained in systems external to TDMS. Integrating the data into a single repository allows for the establishment of relationships between various elements currently stored in separate, isolated information systems.

Recent analysis of the existing TDMS and YMP business processes revealed data are tracked in multiple systems that use manual processes to transfer the data between systems. This often involves retyping the information into a new input screen or application. This manual process lends itself to introducing errors, usually typographical mistakes. The modernization of the systems within the scope of TDMS will help to eliminate many of the errors. Once accomplished, the data can then be integrated with data from external systems into a DW. The warehouse building process includes checks that detect inconsistent and duplicate data, establish correct relationships, and enforce business rules to facilitate the data exchange among different platforms.

2.0 WAREHOUSE SOLUTIONS

The TDMS is comprised of several different applications and data stores. The redesign of the existing TDMS will help to integrate a large portion of the technical data gathered, maintained, and used to support the LA of Yucca Mountain. A high-level diagram of the existing TDMS architecture is shown in Figure 1.

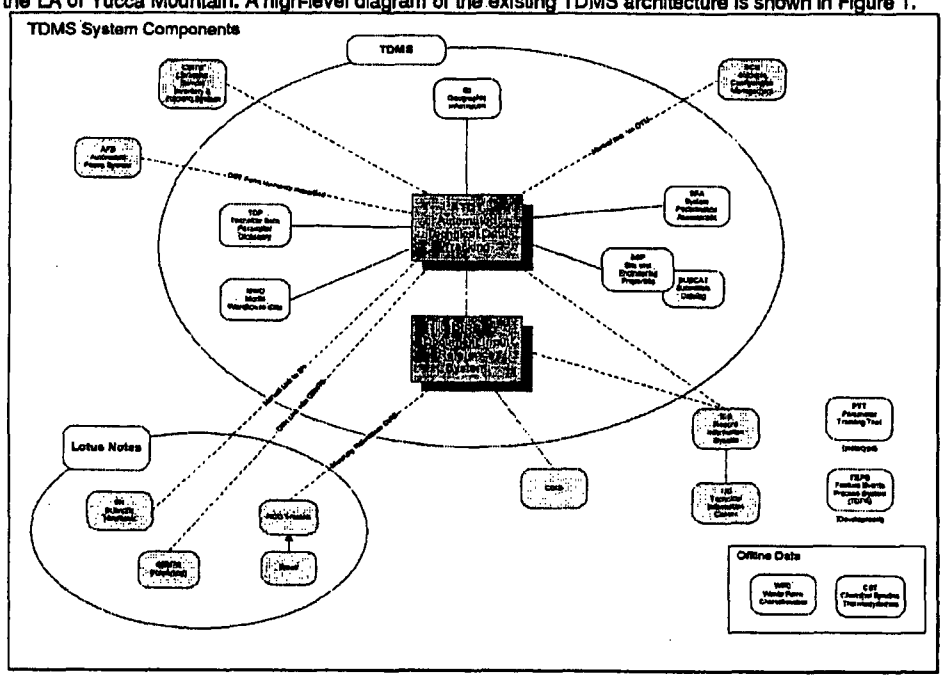


Figure B1. TDMS Context Diagram

Several key components needed for the entire YMP reside outside the scope of the proposed replacement for TDMS. Key information from Lotus Notes, CDIS, RIS, TIC, SCM, and CSITS is crucial to tracing data from acquisition to scientific conclusion. Currently, there are manual links and processes established to track pieces of data both internally and externally to TDMS specific systems. DIRS requires information from numerous supporting systems to support the reference verification process as well as the document reference management process. For instance, the DIRS administrator currently must manually check the status of documents in CDIS to see if a document that has also been tracked or referenced in DIRS has become final. Once the status in CDIS becomes final, the corresponding document in DIRS must also reflect that change.

A DW implementation will help keep these systems synchronized automatically. Furthermore, it would allow for information interchange between the systems with minimal impact on external systems. Processes can be established to update the DIRS system automatically or allow the administrator to review and validate the data prior to the status of information being updated.

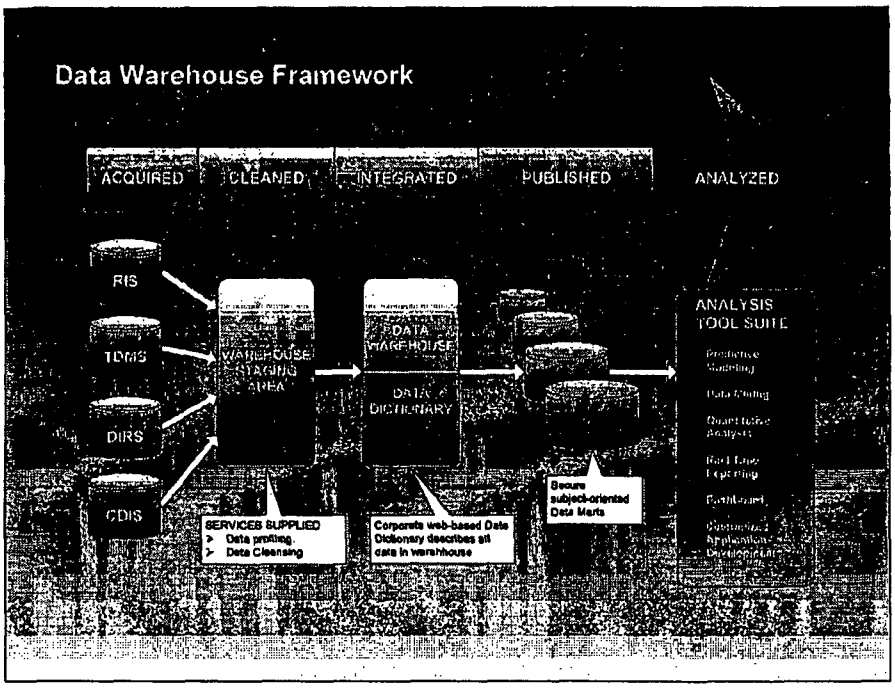


Figure B2. YMP Data Warehouse Framework

Figure B2 shows a logical view of the proposed DW framework for the YMP. The warehouse acquires data from numerous operational source systems (such as TDMS, DIRS, RIS, etc.) and loads it into the warehouse staging area. This provides a place for analysts to analyze the data and assess its quality and completeness as well as clean the data before it is loaded into the core area of the warehouse. The DW core is the principal area where data are stored, and is also the area

where data from disparate sources are integrated. A customized subset of data from the core warehouse is then published to a data mart that the business intelligence tools use to analyze and report data in a meaningful way.

There are many solutions that a DW could provide in supporting the license defense stage of the LA for the Yucca Mountain repository. A few of these are listed below:

- **Integrating disparate data into a single repository.** The data that currently reside in various disparate systems outside of TDMS (such as Lotus Notes, CDIS, RIS, TIC, SCM, and CSITS) can be integrated into a central repository. This creates a more complete picture of the YMP. Regulatory documents, parameter data, source data, developed data, and documentation will be combined and linked to provide a total data life-cycle process. This enables the traceability of data being reused back to its original source. Integrating data in a DW repository from all of the associated YMP systems will enable users at all organizational levels to analyze, correlate, and report information in a meaningful way that could not be performed practically otherwise. History tracking allows users to see what items change and the impact of those changes. Users can perform trends analyses and predict possible future events based on the patterns seen in the data collected. Finally, analytical tools (such as data and text mining tools) will have access to a complete set of integrated data on which to perform their analysis.

- **Reducing manual processes.** Many of the manual processes currently used in TDMS can be eliminated with a DW. Information about data change can automatically be propagated to other systems downstream as updates occur. Managers and administrators will be able to see how changes will affect dependent data items. IRAN notifications can be sent automatically when changes to technical data impact other data sets. The integration of the data will allow for the automatic retrieval of Technical Parameters, Test Data ID and Sample Management Facility ID numbers from CSITS, and Software Tracking numbers from SCM to be linked into TDMS and DIRS records. TIC and Catalog numbers and copyright status information can also be automatically fed into DIRS. Accession or Package ID information and status changes in RIS records can be updated in DIRS automatically.

- **Enhancing data quality and integrity.** The integrity and quality of essential YMP scientific and engineering data would be enhanced since part of the warehouse building process includes checks to detect inconsistent and duplicate data. The information can then be distributed to the people who maintain the operational system for correction. Furthermore, discrepancy reports can be generated during loading and update processing to identify inconsistencies in data values between different systems. Reference information can easily be verified and validated. Automated data feeds between systems will reduce data inconsistencies caused by re-entering the same data.

- **Enabling queries that cut across different databases.** The need to have stronger mechanisms within TDMS to pull data from various data sources can be addressed via a DW. Through a central data repository, queries and custom reports against various systems (such as CDIS, RIS, TIC, etc.) could be easily built and maintained, thus decreasing the workload on operational systems. In a DW, data can be imported in virtually any format and the manipulation can be automated so that data in previously incompatible systems can be compared and reported on.

- **Providing a platform for analytical queries and integrated reporting.** Generic reports can be written to extract all related data based upon user input. For instance, a query could be created to allow an authorized user to report on all associated data items (inputs) and resulting data sets (outputs) for a given AMR. Another example is the ability to list AMRs that would be impacted by the changes in separate AMR. An integrated DW would enable reporting on the DTNs currently being used as well as those not referenced, including their qualification status. Integrating FEPs data in the warehouse will enable identification of regulatory requirements required for the performance assessment analysis of Yucca Mountain. Acquired data, gathered in accordance with FEPs, can then be input into analysis models with the results captured in an AMR.

12.0 RISKS

- Systems outside of TDMS currently reside on separate networks.
- Connectivity between networks could be problematic. Sandia will not have control over those networks or resources. This risk can be mitigated by establishing strong support agreements among interested parties.
- Cleanliness of the data is crucial to building an effective data warehouse. Subject Matter Experts (SMEs) will need to be available to review, verify, and validate the data. Corrections should be handled at the source system.

13.0 ASSUMPTIONS

To build the integrated DW, information must be electronic and suitably formatted. Data in proprietary systems would need to be extracted into a suitable format (such as fixed record length or delimited formatted files). Data that cannot be transformed into a suitable format cannot be integrated.

Hardware and network connectivity would be established and maintained by the YMP Information Technology (IT) Infrastructure. Quotes are listed below for personnel, hardware, and software to develop a data warehouse for YMP by Sandia. These costs do not include long-term sustainment after the initial deployment. Connectivity between separate networks would be established and maintained via a service level agreement or a memorandum of understanding. SMEs must be available to answer and resolve questions and data discrepancies.

APPENDIX C: POTENTIAL DART CONTRIBUTION TO TDMS

1.0 INTRODUCTION

The DART project encompasses a number of tools and technologies. Not all DART offerings are relevant to the Yucca Mountain Project (YMP) Technical Data Management System (TDMS). The portion of the DART toolset that provides document and data management services is called the DART Simulation Data Management Toolset, or DART SDM Tools. It is the DART SDM Toolset that is most likely to be of interest to a TDMS redesign.

One of the primary strengths of the DART environment is its data management model. Included in this model are several features of potential interest to TDMS:

- a mechanism to group data into logical sets called artifacts, and to organize artifacts into a structured project
- the ability to keep a history of data changes, including attributes such as who made a change and when
- the ability to annotate data with time- and user-stamped comments
- the ability to track input-to-output dependencies between artifacts, including which specific versions of artifacts are involved in the relationship
- the ability to restrict access to data to specific authenticated users, groups, or roles
- the ability to extend the existing data model to accommodate new types of data

The DART data model is implemented within the DART SDM Toolset. A more complete description of the DART data model can be found in *An Introductory Guide to the DART Environment*, currently available as Web FileShare document 448860.

14.0 DART SDM ARCHITECTURE

The SDM Toolset has several software components: Oracle, eMatrix, the ESAW server, the APC Server, and the APC Client. The SDM Toolset utilizes three main hardware resources: the SDM Server, a corporate SAN, and the user's local workstation.

SDM uses an Oracle database as its foundation. The Oracle database is hosted on the SDM server.

A commercial application called eMatrix sits on top of the Oracle database, enabling file management capabilities. You can think of eMatrix (at least how we use it) as a souped-up version of the source code management tool CVS. Like CVS, it maintains a history of each file, a comment about each file revision, and attributes like time stamps. In addition, eMatrix organizes files into groups, allows the creation of relationships between file groups using links, and provides a means to apply generic attributes to file, file group, and link objects.

eMatrix does not store files on the SDM Server. Instead, it has been configured to store files on a corporate SAN. The files are stored on the SAN, while attributes are stored in the Oracle database on the SDM server.

eMatrix has its own data model, which it exposes through the eMatrix API. The DART SDM Tools have their own data model as well, one of which is more suited for management of analysis project data. By defining its own data model instead of using the eMatrix data model directly, DART has been able to refine the data model as needed. Using its own data model also introduces a layer of abstraction between DART and eMatrix, making it possible to use something other than eMatrix in the future.

There are two SDM applications that access eMatrix directly, the ESAW server and the APC server. These two server applications can be thought of as translating between the DART data model and the eMatrix data model. The two applications share a common understanding of how the DART model is represented in eMatrix. Data created in one server application is also visible in the other server application.

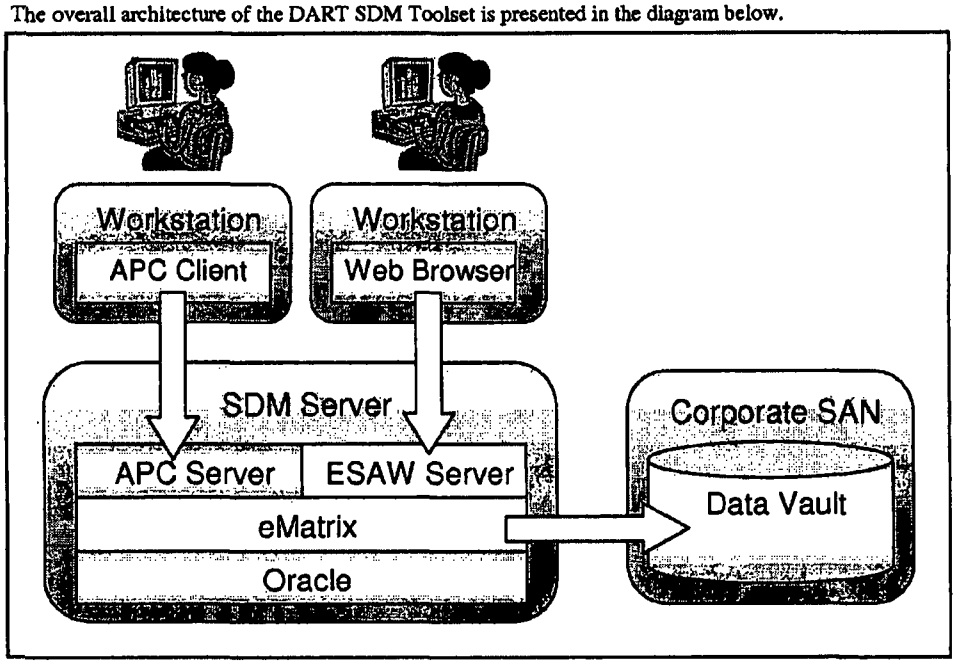
ESAW is primarily a project management tool. The ESAW Server provides a web-based interface. Using a web browser, users can create projects, set up project teams, define project deliverables, and associate specific file groups with a deliverable. ESAW also allows a workflow to be defined for a project. The workflow capabilities in ESAW are oriented toward the requirements of the weapons programs.

The APC Server provides data management services such as uploading and downloading files, and establishing or querying history and pedigree information. The APC Server provides an API which is accessible via the HTTPS protocol.

The APC Client is a rich client installed on the user's local workstation. The APC Client interacts with the APC Server through its API, providing an interface for users to view and modify project data.

Note that any component of the DART SDM Toolset architecture can be replaced with an alternative. For example, an application other than the APC Client could be used to communicate with the APC Server using the APC Server API over HTTPS. The APC Server can be made to work with an alternative to eMatrix by providing a software component called a database adapter.

The overall architecture of the DART SDM Toolset is presented in the diagram below.



15.0 ABILITY TO SATISFY TDMS REQUIREMENTS

Requirements related to business rules would typically not be addressed by the DART SDM tools, but would be handled by other software tools that interact with the DART SDM tools. Non-functional requirements are also not typically addressed by the DART SDM tools. The non-functional requirement to control privileges based on roles and groups is satisfied by the DART SDM tools.

Functional requirements are addressed by the DART SDM Toolset when used in conjunction with other software. The DART data management core is intentionally both generic and extensible. The flexibility provided by its generic data management model makes it possible for DART to satisfy nearly all of the TDMS functional requirements, but only after being customized and integrated with other tools and technologies.

Within the DART data model, data is grouped into generic data sets and pedigree relationships between data sets are represented as generic directional links. The generic nature of the data management capabilities allows the DART system to be utilized in a wide variety of settings with many different types of data.

The move from a generic system to a specific system is achieved through built-in extensibility features. The DART system includes a number of built-in extensibility mechanisms that allow projects to define specific attributes that can be stored and searched.

To utilize DART, Yucca Mountain-specific data schemas would need to be defined, and tools that work with Yucca Mountain data would need to be developed. Appropriate user interfaces to interact with data would also need to be developed, as the user interface provided by the APC client and ESAW browser interface are not ideal for the TDMS environment. This could be achieved by modifying the existing APC client (it is built within a customizable Eclipse framework) or by replacing the desktop APC client with a new application that utilizes the APC Server API to access the data repository.

16.0 COSTS

There are three main costs incurred with a DART-based TDMS system:

- Oracle and eMatrix licensing costs;
- Development of data schemas that are couched within the DART schema extensibility framework;
- Development of software tools that interact with the DART server applications while imposing the TDMS business rule requirements.

Oracle and eMatrix licensing costs have not yet been determined.

DART can provide data storage, history, and pedigree capabilities for a very low cost. The DART SDM Toolset already provides these capabilities, and would be provided to YMP for free or nearly free. The expense comes when the rest of the system is built around DART.

YMP-specific XML Schemas for data representation would need to be developed. The cost of developing these schemas within the DART data model would be similar to the cost of developing those schemas from the ground up (this is probably true for any alternative to DART as well).

It will also be necessary to develop one or more applications that implement business rules and provide an interface to TDMS. This could be done by either replacing or extending the APC Client. The APC Client can be extended in a cost effective manner by writing Eclipse plug-ins (see www.Eclipse.org). This approach can be significantly less expensive than writing a stand-alone application. Several DART plug-ins have already been developed to view, validate, or edit specific types of data. It is also possible to build a completely new client from the ground up. The new client would communicate with the APC

Server using HTTPS. This option is almost always more expensive than writing a plug-in, but comparable to providing the same functionality in a solution built from the ground up.

17.0 RECOMMENDATIONS

It is recommended that the DART SDM Toolset be more closely examined when preliminary design alternatives are being considered. DART provides a number of key features, particularly data storage, data history, data annotation, and data pedigree capabilities. However, the DART system is not a complete TDMS solution right out of the box. It is a component of a larger architecture which requires custom development to support the appropriate business rules and the desired user experience.

XXX

APPENDIX D: POTENTIAL SDDB CONTRIBUTION TO TDMS GRAPHICAL REPRESENTATION

Part Groups

R	C		Part Group	Description	Type	Last Modified By	Date Modified	Reviewed By	Date Reviewed	Certified By	Date Certified
✓		☐	ACTUATORS/SQUIBS(U)	Actuators/Squibs, Unclassified	UNKNOWN	jimasci	03/05/2007 09:58:04				
✓		☐	CASES (U)	Cases - Unclassified	Other	jimasci	03/05/2007 10:06:15	thpeter	03/05/2007 10:07:31	mvsoder	03/05/2007 10:08:09
✓			MECHANICAL HARDWARE(U)	O-rings, brackets, bolts, cover plates, rings, etc., Unclassified	DSD	jimasci	03/05/2007 09:59:45	thpeter	03/05/2007 10:07:31	mvsoder	03/05/2007 10:08:09

Save | Delete Selected Item | Assign | Reset

QA/QC Performed on Characterized Parts and Locked

Stockpile Dismantlement 1.6.0_15 | Enter Data | Review Data | View Data | Maintenance | Help

Home / Disposition Tree Part

For Selected Branch (310999-00):

Last Modified By: weblogic

Date Modified: 2007-03-05 10:34:48.0

310999-00 Hazard Rollup Tree

- [redacted]
  - 1 ea, 310314-00, 0 Synonyms, Part Summary, Characterize Part INSTRUMENT PANEL,
  - either, 1 ea, 310317-00, 0 Synonyms, Part Summary, Characterize Part Bracket, Support,
  - [redacted]
  - 1 ea, 310332-02, 0 Synonyms, Part Summary, Characterize Part Component Level Widget CABLE,
    - [redacted]

Color Legend:	Not Started	[redacted]	Partially Complete	Needs Certifying	[redacted]	[redacted]	EBOM Import	[redacted]

Disposition tree for parts

APPENDIX E: COTS ANALYSIS

Formatted: Indent: Left: -0.35", Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0.25" + Tab after: 0.6" + Indent at: 0.6"

1.0 PURPOSE

The purpose of this artifact is to summarize the resulting requirements of the recently completed Data Management System (TDMS) analysis phase and propose a path forward prior to the design phase of the project. The analysis phase focused on gathering functional and system requirements for a subset of existing applications developed to manage the storage, review, approval, tracking, and traceability of technical data generated to support the licensing effort for the Yucca Mountain Project. From the requirements gathered, this document will list potential solutions including commercial off-the-shelf (COTS) alternatives.

2.0 PROBLEM

Although TDM is comprised of primarily two parts, TDMS and DIRS, it relies heavily on and is tightly coupled with several external systems. This coupling, however, is not direct. The connectivity between the systems is made manually. Identifiers that join common document records between the systems rely on human interaction (cut and paste generally). At a very high level, these separate systems provide similar functionality and store duplicate meta-data in their databases. The primary purpose of these systems relates to data management and workflow. The combination of disjointed system integration, replicated meta-data between systems, and a lack of common data management processes among disciplines have made the management (tracking and traceability) of YMP related documents difficult and, in some cases, erroneous.

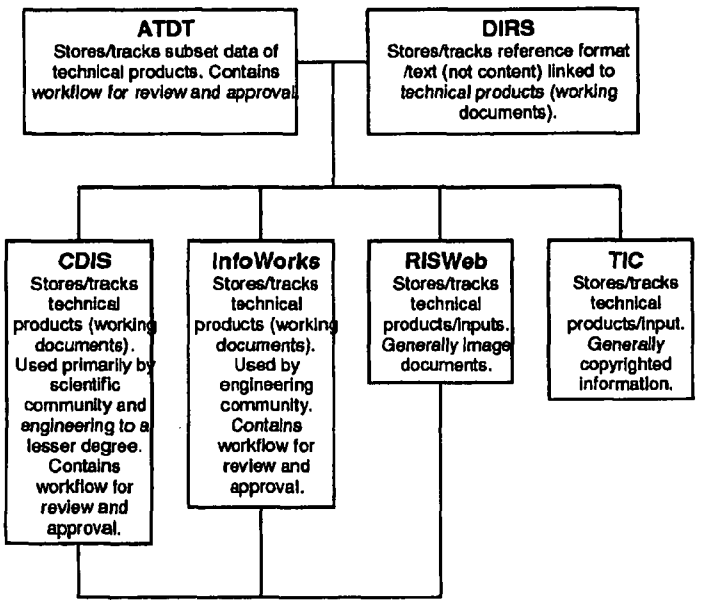
2.1 Background

Process

Considering OCRWM-wide sources of input, differing processes and/or procedures exist for managing data. Not only are different systems used for data management, but different data/document identifier formats are used. In both cases the identifiers are not system generated, and are instead free-form entered by system users. Meaning is assigned to document identifiers, and in many cases, users who generate these identifiers are unsure as to how the identifier should be formatted, creating the need to update identifiers frequently. Also, the engineering community utilizes a process by which they can create "placeholder" identifiers that do not have a document linked to them. Apparently, the meaning assigned to identifiers has no useful purpose nor is it DOE mandated, but it is a substantial source of inefficiency as duplicate identifiers can be created and the identifier format is not widely understood by users.

Primary Components of Existing Architecture

Shown below are the primary components of TDMS and how they relate to external components. Replicated functionality can be seen in almost every component. Note that the bottom four components are all document management systems. Also note that a technical product record can span across as much as 5 of these components linked by a common identifier that is manually entered into each system along with replicated meta-data such as author, title, etc.



3.0 METHODOLOGY

The steps followed to determine the possible solution options and recommend the path forward were:

- Understand the problem,
- Gather requirements for the system, and
- Investigate existing capabilities.

3.1 Understand the problem

It was critical to understand the history and issues that led to the current state of managing classified documents/media. A great deal of time was spent meeting with stakeholders. This includes stakeholders:

- • involved in previous research efforts
- • administering the current systems
- • defining or responsible for implementing policy

- using the current systems, and
- with similar interests or goals.

3.2 Gather Requirements

In order to determine the scope of the project, the business owners and business processes to be enabled were identified. Once these were established, the set of requirements necessary to enable the processes were gathered. In the case where business owners and end-users had conflicting requirements, business owners' requirements took precedence. Validation of requirements was done through iterative meetings with the business owners. Upon completion of validation, the requirements document will be provided to all other stakeholders.

3.3 Investigate Capabilities

Current content/data management capabilities were identified to determine general available functionality. Capabilities include both in-house developed solutions and COTS options. Many of the options listed below relate to document management/workflow which can also handle the storage/management of more discrete data. The options surveyed were:

WebPE

WebPE was presented to a segment of BSC last year as a possible replacement for TDMS. WebPE, Inc. is described as being a specialist in environmental software and portal technology.
Listed capabilities include:

- *Advanced workflow support such as version control, approval process, event management and alerts.*
- "Need to know" access and deployment for large, complex groups
- Dynamic reports providing instant, relevant, and accurate results
- Includes both documents and data
- Portal integration
- Tools to store and technical libraries to work with discrete data including:
  - Analytical Methods
  - Chemistry
  - Field Measurements
- Geography
  - Hydrogeology
  - References
  - Sampling
  - Thresholds / Criteria
  - Well Construction

SharePoint

Microsoft's SharePoint appears to include an Enterprise Content Management (ECM) component. Capabilities include:

- Document Management: check in/checkout, versioning, content types, metadata.

Formatted: Bulleted + Level: 1 + Aligned at: 0.25" + Tab after: 0.5" + Indent at: 0.5", Tabs: Not at 1.9"

- Workflow: review/approval, light-weight process customization.
- Policy and Compliance: Records Management, expiration, IRM protection, auditing.

SharePoint now includes workflow capabilities built on Windows Workflow Foundation that are integrated with the document management repository and forms capabilities.

Documentum

Documentum is a leading ECM product owned by EMC software. It has the ability to record the complete history of movement of documents and custody of documents, including the daily activities of creating, transferring, checking in and out, and destroying documents. It can support inventory and audit reporting needs, has bar-code capability, and provides security at the functional level. The strength and focus of the Documentum products is for management of electronic documents; however, it does have capabilities for managing paper documents as well. Sandia has purchased a 10,000 user license for the records management and content management components to be used for the Electronic Archiving project with the Records Management group.

Stellent

Stellent is a web-based content management product. As with most current ECM products, its focus and strength is on managing access to electronic documents. It has good capabilities in tracking electronic access to documents, but has significant shortcomings in tracking movement of physical documents. Although intended as a version control mechanism, it does have check-in and out capabilities that can be used to track access to documents. It also tracks view accesses. Stellent is the product used for the IES Web FileShare (WFS) application. It makes use of the corporate metagroups for its access control and resides on both the SRN and the SCN.

Oracle

Oracle has a product called Oracle Files, which a 2004 Gartner report described as providing "light weight file management". A new product, Oracle Files 10g, will focus on enterprise content management and workflow capabilities. Both products are applicable to electronic data content. SNL currently does not own Oracle Files or Files 10g.

4.0 RECOMMENDATION FOR SOLUTION OPTIONS AND PATH FORWARD

1.1 SOLUTION OPTIONS

The following options have been listed in descending order from maximum to minimum consolidation scenarios. Note: COTS products were evaluated for high level functionality to meet the needs of TDMS and related systems. Where COTS products are identified as potential solutions, further investigation would be required to assure advertised/claimed capabilities are compatible with TDMS at the lowest levels.

Complete Consolidation

Replace all disjointed systems having similar functionality with one product whether in-house or COTS solution. This would include combining TDMS, DIRS, CDIS, InfoWorks, RISWeb, and TIC functionality into one system. From the limited analysis performed on CDIS, InfoWorks, RISWeb, and TIC, it appears that these systems may have the potential to be housed under one system, but further analysis would need to be performed to validate this assumption. SharePoint may be the obvious COTS solution

for this option since it will be implemented for YMP, but WebPE or Documentum may provide more robust functionality. Pros include optimum traceability, tracking, and data integrity. Cons include highest cost, longest time to implement, extensive staff re-training, and greatest data migration efforts.

Limited Consolidation

Replace TDMS and DIRS with WebPE to interface directly with Data Warehouse. Write new interfaces so remaining systems (CDIS, InfoWorks, and TIC, RISWeb) have direct access to each other and WebPE. This will eliminate the manual connectivity that exists now. Pros include enhanced traceability, tracking, and data integrity. Cons include high cost, limited staff re-training, and greatest development efforts.

Integration In Lieu of Consolidation

Maintain the existing suite of applications related to TDMS. Provide Data Warehouse to contain replicated common data from CDIS, InfoWorks, RISWeb, and TIC. Rewrite TDMS and DIRS in-house to interface directly with the Data Warehouse eliminating the manual connectivity that exists now. Pros include enhanced traceability, tracking, data integrity, and little or no staff re-training. Cons include high development efforts and greater reliance on batch processing.

1.2 RECOMMENDATION

In light of the limited time remaining for the next license application scheduled in 2008, our recommendation is to enhance system reliability, traceability, and data integrity by proceeding with the integration in lieu of the consolidation option. This option provides the greatest possibility of delivering an enhanced and reliable system in time for use during the LA process, with little or no staff re-training to speak of. Design and development is scheduled to begin in Oct. of 2007, making the complete consolidation option less feasible, as the extended analysis required, data migration and implementation efforts would most likely require more time than would be available before the LA start date. It is not recommended to select the complete consolidation option. While the limited consolidation option may be feasible, it is not recommended due to the customization efforts required to interface with multiple external system protocols. Furthermore, a COTS solution/integration at this time is not recommended, due to extensive staff retraining and potentially intense customization/integration efforts. The limited time available before the next LA would not allow adequate time to implement a COTS solution successfully.

APPENDIX F: TDMS ISSUES

The following table provides a detailed listing of TDMS issues and gaps that have been identified to date.

Overall System Issues		
Issue Identified	**Recommendation**	**Map to BCS TDMS CPR**
Current system lacks implemented controls (i.e., business rules) and verification checks. They are implemented by humans causing lack of data integrity.	Desired state should incorporate business rules in computer system code and automate verification checks as much as possible. (Need to identify specific BRs and verification checks)	
Many input processes are manual that could easily be automated. Currently TDM staff employs error-prone, low-level computer commands and techniques normally used by technical computer database administrators and programmers.	This problem should be eliminated with the incorporation of new technology (e.g., automation of many processes). (need to identify specific problems)	
User read/write privileges on TDMS databases are not controlled by Role and Group.	User access controls using role/group can be implemented in the "to-be" system. (Will need to identify role-based privileges for those that will still exist after process redesign.)	
50% of the datasets are not active, which affects performance.	Create a separate data mart for unused data in the warehouse? (Will need to identify unused data)	
There is duplicative data in the various databases due to 1) change management accomplished using DTN's and 2) multiple input points that are largely independent of each other, so the system is unable to recognize same data as duplicate.	Accomplish change management electronically using tool. Actual cleaning of data should happen during data migration. Cognitive Sciences group might be able to help with tools being developed for short term. Single database and process change should also help. (For existing duplicates, need to identify types of duplicity that exists during analysis.)	
Issuing of IRAN manually to affected organization causes delays. Also, finding correct and complete list of affected organization/owners is a problem.	Process needs to be further analyzed to identify parts of this process that could be automated or changed. Some aspects might fall outside our current scope.	
Vendor data also kwon as QSL data (Qualified Supply List) are stored in TIC. When this QSL data changes, there is no provision for updating the changes into TIC nor is IRAN process in place for notifying the users of the old version to do an impact analysis	Gap in the Scientific Investigation process. Need to redesign the process.	

Overall System Issues		
Issue Identified	**Recommendation**	**Map to BCS TDMS CPR**
Currently there is a manual link between SN Registrar and ATDT. SN Registrar is currently in BSC space. Should it become part of TDMS target system or stay as is?	Provide integration capability between TDMS and SN Registrar? How does this fit into the current architecture and what should the future recommendation be?	
The off-line systems may have to be re-examined to ensure all the relevant data has been successfully transferred to the active TDMS systems.		
Some processes are entirely manual requiring a person to verify and validate data items.		

Analysis of Current TDMS Submittal and Incorporation of Data to TDMS (Issue types can be traced directly to sections in TST-PRO-001 that can be found at https://wfsprod01.sandia.gov/intradoc-cgi/idc_cgi_isapi.dll?IdcService=GET_SEARCH_RESULTS&QueryText=dDocName=WFS421737.		
Issue	**Recommendation**	**Map to TST-PRO-001**
Process of preparing data package in accordance with the documents SCI-PRO-005, SCI-PRO-006 and TST-PRO-003 results in inconsistency due to originators' unfamiliarity with these documents. Data coordinator guides the originator.	With the help of machine learning can we provide originators guidance in making more consistent decisions? (according to Mike Jaeger documents listed above have enough guidance that this approach could be possible) (someone with DTM knowledge will have to validate this)	6.1.1 A
Data Tracking Number Request Sheet (DTNRS) is manually retrieved from Automated Form System and filled out by the originator as part of preparing data package.	Provide capability for originator to directly enter all information needed to facilitate the identification and submittal of data and technical information to TDMS?	
Identifying TDMS Database in which submittal is to be stored is a required field in current DTNRS and is a cause for inconsistencies and confusion.	Eliminate many of the subsystems for the "to-be" system, potentially one database? System could help the originator select between electronic and physical format based on decision criteria currently used by data coordinator? Depending on type of information being submitted, an originator might be required to provide different information.	
Manually tracing supporting data to its applicable SNB's, its page #'s, photographs, etc. causes inconsistencies.	Should be handled by the system instead of managing traceability. However, SNB's are currently stored in RIS, which is out of scope for now. Integration with this system should be further analyzed in future.	6.1.1 B

Analysis of Current TDMS Submittal and Incorporation of Data to TDMS (Issue types can be traced directly to sections in TST-PRO-001 that can be found at https://wfsprod01.sandia.gov/intradoc-cgi/idc_cgi_isapi.dll?IdcService=GET_SEARCH_RESULTS&dQuery1txt=dDocName=WTS421737)		
Issue	**Recommendation**	**Map to TST-PRO-001**
Manually tracing data causes inconsistent (traced to wrong source data) and missing traceability of developed data to its source data.	Should be handled by the system instead of managing traceability manually; thus originators will not be required to provide this information and there will be better consistency.	6.1.1 C
Issues with correct identification of acquired vs. developed data by originator.	System might be able to identify acquired data vs. developed data based on their definition?	6.1.1 D
When a Product Output is deleted out of an AMR and has been used by another as Product Input no process to notify the users	Need to redesign the SIP.	
Data Admin is allowed to change the data into TDMS before replies to IRAN are received. Also, receipt of IRAN replies has open time limit.		
Review of originator data submittal package first by the data coordinator and then again by a reviewer might be unnecessary redundancy in the current process. (May be currently necessary because so much of the process is manual and it helps catch errors.)	By automating some manual processes, we may be able to get by with just one review. For "to-be" system need to analyze the data coordinator role; by incorporating some tasks into the system this role may be eliminated?	6.1.2
Manually assigning DTN's for DATA and Scientific Information causes inconsistencies and missing DTN's, resulting in major problems in finding accurate, timely and consistent information in TDMS that is required to support LA.	The purposes for creating DTN's are traceability, unique identification, and searches; this could be done by the system instead of current manual process. Need for creating and tacking DTN will be eliminated.	6.1.2 C
Completing TDIF hard copy	Notes in 4.1.5 also apply here.	6.2 and 6.2.2
	Changes to completed TDIF could be automated and this step might become unnecessary? Need to further understand the process.	6.3
Creating new DTN when data or scientific information with existing DTN changes results in problems with duplicates, missing DTNs and inconsistencies.	Instead of creating new DTN change history could be automatically be maintained by the system, thus eliminating existing problems.	6.5

Analysis of Current TDMS Submittal and Incorporation of Data to TDMS (Issue types can be traced directly to sections in TST-PRO-001 that can be found at https://wfsprod01.ismdin.gov/intradoc-cgi/idc_cgi_isapi.dll?IdcService=GET_SEARCH_RESULTS&IdcQueryText=dDocName=WFS421737)		
Issue	**Recommendation**	**Map to TST-PRO-001**
Providing documented evidence of a completed technical review for Un-Q data and then manually making changes once documentation is provided, causes delay in accurately identifying data qualification. Creates unnecessary manual labor causing bottleneck for data coordinator and the originator	This process should be automated to the extent possible. Some parts may be out of scope for this phase since that might require QA process of the origination sites to be integrated to TDMS. However, some aspects of the process could be automated to make transaction easier.	7.1
	DIRS	
Document owner list is not always current .. When the original owner leaves the organization no process that updates the document owner name	Redesign the process	
No integration between CDIS and DIRS. CDIS does not know when document in DIRS is locked and DIRS does not know Doc status changes in CDIS (i.e., document is cancelled in CDIS)	Tight integration between CDIS and DIRS is needed.	
Working Documents are given due date (author provides when he/she expects completion). But in the current system no provision for monitoring due date (may be manual) and or to take further action when due date passes.	Process redesign is required.	

Model Warehouse Issues		
Issue	**Recommendation**	**Map to BCS TDMS CPRs**
Currently users may not have access to the latest versions of software used to create models in the warehouse. In order for them to get the most current version they have to contact SCM. We might want to look at the current procedure and see if the issue can be helped.		
	GIS	
Maps created from the acquired data from the TDMS are currently stored in folders, but they should be scanned and stored in RIS	Process redesign.	

APPENDIX G: Data and Text Mining Analysis

5.0 INTRODUCTION

Data and Text Mining (DATM) has been shown to help humans sift through massive amounts of information and glean relevant information. The primary goal of the Yucca Mountain Project (YMP) is to support congressional authorization, oversight, and regulatory compliance. Beyond structured data, the YMP systems contain terabytes of unstructured text and information. While it would not be possible for a human to analyze every piece of information, DATM techniques can be used to deal with overwhelming amounts of information. Data and Text Mining can help facilitate these information extraction processes and save time, while users can have high confidence in the methods.

The License Support Network (LSN) contains a collection of scientific documents which will be accessed during the license defense process. Since it contains large amounts of unstructured text, and it is unfeasible to sift through this data manually, we propose to use some DATM techniques to help solve problems not solvable by traditional software engineering applications. These techniques can help create a system comprised of largely textual data that can be queried efficiently, and from which accurate and complete information can be obtained.

Sample scenarios of common work that can benefit from DATM are:

- **Keyword Recommendation** - Scientists must choose their own keywords when entering information into the system. It can be time consuming for them to determine which keyword is most appropriate for their document, and it can be an error-prone process.
- **Elimination of Duplicate Information** - Scientists might unknowingly insert duplicate information. In addition, while two sets of data might not be syntactically identical, they might actually be conceptually identical, and it could be very useful to know this.
- **Concept Searching** - Analysts are currently able to search by keywords. Their effectiveness will be improved if they are able to search at a conceptual level.
- **Trace Suggestions** - All of the records that are inserted into the system must be qualified. This qualification is accomplished through the creation of a tree-like structure, the trace, which is comprised of the complete information required to support the validity of the record. Currently the links to the qualifying record have to be manually inserted, either by the scientist when creating the record or by the analyst when doing a trace of qualification links. DATM based tools can speed this process up dramatically, and can also help analysts be certain that there is no information missing from the trace.
- **Question Anticipation** - The defense of the license must occur on a deadline that is too short for humans to do a thorough analysis. Unanticipated questions will undoubtedly arise during the defense process. It could be useful to predict topics of possible questions that might arise so that analysts can make sure the answers can be retrieved, and in a timely manner.

6.0

7.0 DATA AND TEXT MINING RECOMMENDATIONS

We propose to use DATM to address the challenges described above.

Keyword Recommendation - When a scientist enters information into the database, he is required to select a few keywords to describe the data he just entered. Data and Text Mining can help by automatically suggesting candidate keywords that are appropriate to the database and the information being inserted. This helps keep a unified concept of keywords, further minimizing the ambiguity between what two scientists think a keyword means. An example would be "bank"; one scientist could interpret it as a place to store something while another might think of it as an angle. Another benefit would be that the DATM algorithm might include keywords that the scientist might overlook or not think important. This helps to have a more complete "keyword" record of the data. A further benefit is that this will save the scientist's time.

Elimination of Duplicate Information - The system can scan newly entered records and can initiate an automatic search of the database for similar records. This can help eliminate duplicate records as well as helping to boost the reliability of the qualification of the records.

Concept Searching - A DATM tool can help the scientist find the supporting documents the scientist needs more quickly and accurately. An advanced search can be implemented to search by concepts, rather than by keyword. For example, a search of "rock pressure" would be able to bring up records associated with "rock pressure," even if it does not have the phrase "rock pressure" in it. This capability will provide a larger search space and allow the scientist to see an improved representation of related records.

Trace Suggestions - The traceability of data is a major requirement in the Yucca Mountain Project. Sandia's AMR Exploration Tool will dramatically speed up the creation of the traces. This tool could be augmented with an automated DATM application designed to improve the reliability of the trace, through suggestion of possible supporting records that the analyst may have otherwise forgotten, not known about, or not have recognized as being similar. While an analyst is creating a trace, such an automatic DATM application can be running in the background, suggesting additional data that support the current information being traced. This could help prevent surprises that might arise during license defense.

Question Anticipation - The topics of public discussions can be indicators of types of questions that can come up in the congressional inquiry. We propose to automate the analysis of web pages about YMP, categorize them by topic and attempt to extract public opinion. (Similar information about WIPP opinions and license defense can also be so analyzed.) Once the possible topics and questions that reflect public opinion are discovered, they can be used to interrogate the YMP systems to check its integrity.

8.0 RISKS

Data and Text Mining algorithms are not perfect. The DATM techniques described here are designed to aid the human in the decision making process, and there must still be a human in the loop validating the recommendations made by a DATM tool.

The DATM solutions described in this document will be designed to work only on digitized text and data. Hard copies, scanned free-form data, such as lab notebooks, pictures, sound files, and movies are examples of data that cannot be incorporated into this Data and Text Mining recommendation.

9.0 ASSUMPTIONS

The DATM recommendations in this document are made under the assumption that YMP data exists in a Data Warehouse, where a) business rules are defined; b) the data is in a consistent state, and c) there is a single source of truth. Though DATM techniques may help with the migration of data to a Data Warehouse, those recommendations are not included in the scope of this document.

There is also the assumption that the system is connected to the network where the databases reside and access to the data is provided to the DATM tool. Such a system can be a computer that one of the databases or the Data Warehouse resides on. Care needs to be taken to make sure there is enough machine processing capability for all of the concurrent applications. This system has to be powerful enough to run the DATM algorithms in a reasonable amount time.

10.0 EFFORT

There are two implementation options: existing and new in-house solutions or Commercial Off the Shelf (COTS) packages. Some possible solutions would be:

- In-House
  - Use of the DART tool for search by semantic topic
  - Use of 6341's AMR Exploration Tool for automatic search when performing traceability
  - Use of 6341's STANLEY library for extraction of keywords, YMP guided spidering of the web
  - Transition of 4511's framework for Latent Semantic Analysis into a production level product to do semantic searches, keyword extraction and semantic analysis
- COTS
  - Use of SQL Server's SSIS services for Data Mining
  - Use of Matlab's library for Data and Text Mining to implement semantic searches and keyword extraction
  - Statistica, a Data and Text Mining tool for implementing semantic searches and keyword extraction

Each of these choices has pros and cons usually associated with using a COTS package versus an in-house implementation. The amount of effort required can be estimated only after a selection of the appropriate package.

11.0 CONCLUSION

With the aggressive timetable allowed for the Yucca Mountain Project, Data and Text Mining can help accelerate the evaluation of records and traces, and help scientists, analysts, and

lawyers do their jobs more quickly and accurately. There is an incredible amount of data that resides in the YMP systems and manual evaluation of that data could easily take much longer than the time allowed. Data and Text Mining algorithms can help humans do their work more efficiently while picking up information that the humans might miss.