Professor John W. Harris
Spokesman, STAR Collaboration
Wright Structure Laboratory
New Haven, CT xxxxx

Dear Professor Harris,

Enclosed is the report of the committee charged with reviewing the status of the STAR Software and Computing Project. While this report maitook longer to prepare than planned, in most cases where decisions or directions have been taken, these have been consistent with the advice of the Committee. In other cases where directions are still being defined, the guidance within the report is still timely and relevant. The basic conclusion of the Committee is that with proper focus on the essential tasks which must be completed before startup, the STAR Software and Computing Effort can be successful.

I would like to thank the members of the Committee for their efforts on behalf of the STAR Collaboration, and am hopeful the results of this review will provide important guidance for STAR which will help insure the success of our scientific program.

Sincerely,

Timothy Hallman

STAR Deputy Spokesman
BNL STAR Group Leader

Report of the STAR Internal Review of the STAR Software and Computing Project

Executive Summary

On October 20-22, 1997, a STAR internal review was held at Brookhaven National Laboratory by STAR Management to assess the status of the STAR Software and Computing Project. The purpose of this review was to evaluate the requirements, the design, and the ongoing hardware and software implementation of the STAR Software and Computing Project.

With only two years before STAR commences data acquisition, the purpose of the review was to make sure that the software and computing approach (requirements, design, and implementation plan) is correct and that there are sufficient resources in STAR to successfully carry out the required effort and thus satisfy the physics needs of STAR.

Based on an open review, the committee was asked to evaluate the software and computing design to determine whether the physics driven requirements will be met by the design and its expected implementation. The committee was asked to consider the likely progression of physics-driven requirements from day 1 to year 1 and beyond, and to verify that the plan for developing STAR software and computing will result in the appropriate capabilities as a function of time; for pre-data taking simulations and testing of STAR, for turn-on, for one year later, and for beyond when the full capabilities of RHIC and STAR are available. The scope of this evaluation included the physics requirements for both the nuclear physics program (heavy ion and related proton-ion and proton-proton studies) and the spin physics program

The charge to the Committee and the agenda for the review are included in Appendices I-II respectively. The members of the Committee were:

Hank Crawford	UCB/SSL
Tim Hallman (Chair)	BNL
Bob Jacobsen	LBNL
Mike LeVine	BNL
Pere Mato	CERN
Iwona Sakrejda	LBNL
Peter Seyboth	MPI
Thomas Ullrich	Yale
John Womersley	FNAL

Ex-officio:

John Harris	Yale, STAR Spokesman
Jay Marx	LBNL, STAR Project Director
Jim Thomas	LBNL, STAR Deputy Project Director
Richard Jared	LBNL, STAR Head Electrical Engineer

Findings of the Committee

The Committee would like first to thank the members of the STAR Software team for their hard work in preparing for this review and to commend the members of the group on the level of preparation of the presentations which allowed for a productive meeting and a very useful dialogue. The Committee recognizes that in general a lot of effort has been devoted to software, but notes as well that a great deal of work remains to be accomplished in a short amount of time.

The Committee feels that with proper focus on the essential tasks which must be completed before startup, the STAR Software and Computing effort can be successful with regard to meeting the physics driven requirements of the STAR scientific program from day 1, to year 1 and beyond. The committee views the addition of Torre Wenaus as the Project Leader for STAR Software and Computing as a strong positive step towards strengthening the STAR software effort, but feels as well that for the software effort to be successful, further development and utilization of a detailed milestone schedule and a robust software management structure is essential. The filling of several key positions in this structure awaits full assembly of the core software team. This process is under way, and needs to be completed as soon as possible.

For each major subcomponent of the software project (online, offline, software infrastructure, etc.), a milestone chart should be developed and tracked at a low enough level that each element important for completion of the software project is identified individually. Priorities for completion of different components of the software project should be established so that essential components can be completed in order of their priority in the instance of sparse resources. In that case, to accommodate features or capabilities which are essential in the future but not in the near term, hooks should be left in the appropriate places, so that components of the software which can not be completed as first priority tasks can be finished as resources become available. Responsibilities for completion of the components of the milestone schedule should be clearly identified within the various sublevels of the software management structure. When difficulties in meeting milestones are identified, that information should be communicated up the software management chain to the next highest level, so that a coordinated strategy can be applied to managing available resources in the most optimal way. In the event important problems related to resources can not be solved within the software project, STAR management should be asked to assist in finding the necessary support.

With regard to code development, the Committee notes that since Torre Wenaus has become the STAR Project Leader for Software and Computing, a number of new directions (e.g. Objectivity, C^{++}) have been taken. The Committee understands the need for these initiatives is driven by the desire to provide the capability necessary for the STAR physics driven requirements in the future and by a considered view that to accomplish this goal, these changes are essential. The Committee is concerned however, that to be successful, the STAR collaboration must invest fully in these directions. Because of the pace with which strategy is being developed, a concerted effort may be required to fully articulate and disseminate the need for change, as well as the pluses and minuses of the strategies which are being developed.

A factor key to the success of the STAR software and computing effort will be the extent to which the Collaboration invests itself and utilizes the codes and infrastructure which are developed. For that reason, it is essential that the developments being pursued lead in a timely way to software and infrastructure which are stable, robust, and accessible, so that software which is developed will be readily accepted as the code of choice. Due consideration should be given to developing code which requires modest initial investment, with sufficient tutorials and training available to allow users to become productive as quickly as possible.

The Committee notes that the positions of STAR Analysis Coordinator and STAR Online Leader need to be filled immediately. The Online Project has made a significant (much needed) start with the direct participation of the Software Project Leader, but the Committee feels strongly that unless an Online Leader is found soon, other areas of responsibility of the Software Project Leader may be seriously impacted.

With regard to the progress in the Online Software Project, the Committee feels that it is still early days, and this aspect of the project could not be rigorously reviewed at this time. A separate review of the Online Project within a couple of months is recommended.

With regard to the number of platforms which STAR will support, the Committee feels the Collaboration needs to make a considered decision, taking into account the manpower and cost of supporting multiple platforms versus the benefit of enhancing the capability for software and analysis tasks to be performed remotely.

General Concerns, Conclusions, and Recommendations

Simulations

The committee was happy to learn that software for event generation and detector simulation were at a relatively advanced stage. A variety of event generators have been adapted for use in Star by providing a standard interface to GEANT. Also special mechanisms have been added to generate features like QGP bubbles and DCC domains. The detector hit simulation is carried out by the program package gstar. It is the adaptation of the now frozen GEANT 3.21 of CERN for the purposes of Star, using a special advanced GEANT interface (AGI) language to facilitate the definition of the detector parameters. The program g2t translates the detector hit point output from the GEANT internal Zebra format to STAF table format for further processing. Although a number of preliminary detector response simulation modules exist for the TPC, SVT, TOF, and FTPC, significantly more work is needed in this area. For consistency of simulation and reconstruction, the detector position and calibration data should be stored on a unique database to be used throughout the simulation and reconstruction chain. The committee arrived at the following recommendations:

(1) The simulation and reconstruction should be interfaced to the STAR database (probably Objectivity) as a means of ensuring the use of consistent detector parameters and calibration constants throughout the software chain

(2) Since the advanced GEANT interface is a crucial non-standard Star-maintained piece of software, there should be more than one person dedicated to maintenance and user support in this area.

(3) Final definition of the content and format of the STAF raw data tables is needed as soon as possible in order to proceed with further development of the simulation software (embedding) and with interfacing to the reconstruction chain.

(4) Increased involvement of the detector groups is needed in order to provide much of the effort in completing and tuning the fast and slow (detailed) detector response simulation modules. In particular, such modules are missing for the electromagnetic calorimeter. The

TPC modules should be refined and tuned with the cosmic ray test data which is already available.

(5) Functional scripts and documentation on how to use them should be made available for people working on detector response simulations

(6) The simulation of raw event data should be provided as input for the mock data challenge in order to demonstrate the functionality and throughput of the entire software chain

(7) As a means of deriving reconstruction efficiency estimates, an embedding procedure for placing simulated tracks into real data should be provided. This feature will be required when data becomes available in 1999. Embedding of special physics processes into the event generator output is also desirable in order to study the sensitivity to rare processes.

(8) The integration of gstar and STAF is discouraged since the simulation chain presently is providing STAF table output via g2t and effort is more urgently required on other topics. An analysis and optimization of CPU usage in gstar would be useful.

(9) A transition to GEANT 4 is desirable in the long term since it promises a significant gain in computing speed and improved simulation algorithms. Star involvement during the present development phase should continue at a low level in order to ensure that GEANT 4 will be able to handle large event sizes.

(10) An event-by-event catalog with descriptors specifying the event generator characteristics and the detector simulation parameters should be made available in the Star database.

(11) An annotated list of tasks of Star simulations is recommended which should specify timelines, people actually working, and additional effort required for each item. This list will allow a better assessment of the manpower situation within the simulation effort

(12) Large scale CPU resources for simulation will probably be required starting in 1999. CPU capacity for this task is presently planned in RCF only on an “as-available” basis. A strategy regarding alternative possibilities, including NERSC, for satisfying STAR’s CPU requirements should be developed.

Online Software

1) There has been a tremendous amount of progress in the area of Online but a serious difficulty still remains - a leader is needed for the Online effort immediately. The fact that the Online leader is still missing has obliged Torre Wenaus to spend a large fraction of his time engaged in this activity. While the benefits to Online are obvious, if this situation continues the committee feels that the other aspects of the Software project will suffer. If a leader cannot be found for the Online effort in the next two months, STAR management should consider redefining the Online scope so that it can be absorbed into existing STAR efforts.

2) More consultation between the Online Software group and other STAR Project groups (detectors, trigger, DAQ) should take place, along the lines of the August 1997 Online workshop at BNL which was very successful. During these meetings, detailed scenarios for starting, stopping, and configuring runs should be worked out. Special attention should be paid to specialized activities such as calibration and pedestal runs. These meetings

should serve to collect the requirements from the users (sub-detectors experts, shift operators, ...) of the online system. The run control state model and the interface definitions related to the state model should be finalized by the end of the year. To accomplish these goals, it is recommended that at least two more On-line workshops be hosted at BNL, one of them to be held by the end of the year.

3) The Online project has made major advances in a short time, but it is still in a preliminary phase. Detailed design and implementation decisions have not yet been made. As a result, a further review of Online project will be required after the design is more mature, and technology choices have been made following hands-on testing.

4) It is the feeling of these reviewers that the manpower estimate for completion of the Online software is too low. In particular, it is felt that the effort required to evaluate, test, and learn new software technologies has been underestimated.

5) The current manpower effort in Online is too low to allow development of all pieces in parallel. The committee feels that all the projects proposed by the Online group are necessary, but it is recommended that the Online effort be staged, with the highest priority dedicated to the components necessary to permit the most basic operating mode, e.g., a single stream with data going directly to storage, minimal configuration management, and no event server. Additional functionality which is required should be added as resources allow.

6) The Committee believes the simplicity and flexibility of the current Online state model are very good, and that these features should be retained in the final implementation. It is also felt however that the Online group should be encouraged to look at products developed for other experiments (at Fermilab, CERN, CEBAF) to see whether these might be adapted for re-use within the STAR Online model. It is noted, for example, that configuration management, scripting languages, and run control are essential components of every running experiment, and that some advantage might be taken of existing products.

7) The Online group is also encouraged to examine other parts of STAR, (especially TRG, Slow Controls, and DAQ), for architectural elements which might be borrowed or developed in common. In particular, investigating the use of Objectivity in the configuration databases is recommended.

8) Support of Sun Solaris for Run Control is mandatory due to its installed use in DAQ, Trigger, and Controls. Online should seriously and carefully consider the costs associated with supporting a second platform (Windows NT), comparing them with the possible benefits (available tools, cheaper software and hardware, etc.)

9) The network architecture interconnecting various subsystems in the control room and the physics building requires more study. It is recommended that the Online group look carefully at the bandwidth requirements between DAQ and the event server, as well as between the event server and the client workstations.

Offline Software

New Directions

In a number of areas, STAR is taking a rather untested, future oriented direction: for example, Objectivity, C++, Grand Challenge data-mining, and Windows NT. Being forward-looking is good, and the use of these products will probably result in a better system ultimately, for example, on a 5 year time scale; but it may also give rise to problems on the 1999 time scale as HENP experiments are relatively inexperienced in their use. The choice to move in these directions is therefore not a simple one and it should be made certain that the Collaboration as a whole has bought into this strategic orientation. The C++ migration strategy, user motivated rather than coerced, looks good.

Manpower

Manpower is a huge issue, and is recognized to be so. As well as filling the open positions in the core team, the Committee suggests:

1) prioritizing efforts, e.g. not working on software for unfunded detectors and deferring work on detectors to be installed later.

2) ensuring visibility and rewards within the collaboration for those working on software (talks, names on org charts, mention in DOE reviews...)

3) making the environment easier to work in (STAF) for new people

4) increasing some of the 0.1 FTE people to 0.5 FTE's: rather than adding new names; this preserves investment in training and saves using experts to bring novices up to speed.

Management

The Committee feels that Torre Wenaus has made great progress in organizing the STAR Software Project in the few months since he joined STAR and that he should be commended on this.

Further effort is needed, both within the STAR software effort and by STAR Management, to fill the missing posts in the management structure, and to make sure that the management structure is truly effective. Managers should delegate tasks, and both problems and progress should be reported at regular meetings. Responsibility and tracking should follow an established management chain. Higher-level managers need to be careful not to allow lower-level tasks to interfere with their effectiveness.

Scheduling

The establishment of a Microsoft Project Schedule is an excellent start. Further work is clearly needed to fill in more detail: many more intermediate milestones (besides the Mock Data Challenges) are needed in order to know if subtasks are ahead, behind, or on-schedule, and to help in resource allocation and prioritization.

Reconstruction Code

The Committee was pleased to see that C++ and FORTRAN modules exist and appear to work together. The work done on specifying the design and requirements of the reconstruction code is good, as well as the efforts of the task force directed by Peter Jacobs to identify physics-driven requirements. (It is strongly urged that the work of the task force be completed quickly). Most areas of reconstruction have been started and work has begun to develop a strategy as to how to defer some non-baseline efforts. The manpower estimates are welcome, and look reasonable, although there is no way to calibrate them against past experience in the STAR software effort. (It is felt incidentally that the 18% "overhead" explicitly included as part of these estimates is probably unnecessary, as this is likely implicit in the estimated times).

The Committee feels strongly that tasks within the area of reconstruction should be prioritized and that intermediate milestones should be created (and agreed with the managers of the subdetector-level boxes) so that manpower can be directed between tasks on the basis of their relative state of priority and advancement.

There appears to be two software efforts which have been orphaned: the EMC calorimeter reconstruction, with only 0.6 FTE, and the global reconstruction. Despite the fact that the global reconstruction effort is central to the STAR physics program, it is seriously understaffed (1.0 FTE). It is noted that one contributing factor to the latter problem is that the global reconstruction effort is by nature not tied to any particular detector, whereas other areas of the reconstruction software appear to be able to draw detector-based manpower (e.g. through institutional responsibilities or MOU's). Some effort to formalize institutional responsibility for global reconstruction software may therefore be needed.

A person in the position of Physics Analysis Coordinator is clearly needed. The person in this position would be instrumental in giving input to the reconstruction effort (for example, in terms of prioritization).

Quality and Training

It is the observation of the Committee that design and coding procedures do exist but do not seem to be rigorously followed. These procedures call for a design review before coding begins; this is viewed as a good idea. It is recommended that this means of quality assurance not be limited to interface specification but that the functionality and dynamic operation of proposed modules be included as well. Ongoing guidance and oversight by a review panel is being tried in D0 as an alternative to code walk throughs: it is less confrontational and spreads knowledge of how the code operates.

The state of documentation appears poor and should be improved. There is no development environment, CASE tools, or code-checkers available.

We suggest that the STAF workshops evolve into training sessions for STAR software developers.

If C++ is to play an important role in STAR in the future, then training in C++ will also be needed.

Additional comments on STAF

Analysis Framework

STAR has chosen STAF as its analysis framework. It is a non-OO, table-based package which is intended to ease the reconstruction and analysis of STAR data. STAF provides an interface to user written code in FORTRAN, C, and C++. Recently, PHENIX has also decided to adapt STAF, which dramatically increases the user community by more than a factor of two. This decision however appears to have no direct impact on the responsibility of STAR to support and maintain the STAF package.

The Committee notes that within a large fraction of the STAR Collaboration, STAF has a bad reputation, and is considered hard to learn and cumbersome to use. Whether this is due to the lack of documentation and training and/or reflects too complex a design is not clear. STAF was and is to the present day written and maintained by a single person, who has essentially no full time commitment to STAR.

The Committee feels that this is a critical situation. It is recommended that the responsibility for STAF should be taken over by 1-2 dedicated persons, either from the Collaboration itself, or in a common effort with computing personal from PHENIX and/or the RCF. No further functionality to the framework should be added unless explicitly requested by the user community. The existing code should be improved in terms of reliability, ease of use, and documentation.

Software Infrastructure

For each major subcomponent of the software infrastructure, a milestone chart should be developed and tracked at a low enough level that each element important for completion of the software project is identified individually. Priorities for completion of different components of the software project should be established so that essential components can be completed in order of their priority in the instance of sparse resources. In that case, to accommodate features or capabilities which are essential in the future but not in the near term, hooks should be left in the appropriate places, so that components of the software which can not be completed as first priority tasks can be finished as resources become available. Responsibilities for completion of the components of the milestone schedule should be clearly identified within the various sublevels of the software management structure.

A factor key to the success of the STAR software and computing effort will be the extent to which the Collaboration invests itself and utilizes the codes and infrastructure which are developed. For that reason, it is essential that the developments being pursued lead in a timely way to software and infrastructure which are stable, robust, and accessible, so that software which is developed will be readily accepted as the code of choice. An attempt should be made to keep the Collaboration informed about infrastructure changes anticipated in new releases.

With regard to the number of platforms which STAR will support, the Committee feels the Collaboration needs to make a considered decision, taking into account the manpower and cost of supporting multiple platforms versus the benefit of enhancing the capability for software and analysis tasks to be performed remotely. An optimal plan will probably involve endorsing a small number of select platforms, and discouraging the use of others.

The Committee feels that the raw data and table formats should be defined soon. Also, the deliverables to be expected from the Grand Challenge effort should be agreed up soon between NERSC, RCF, and STAR. With regard to networking it is felt that the accessibility to BNL should be at least OC12.

The Committee feels some effort should be given to insuring against legacy code which becomes orphaned if a given individual leaves the STAR project. One way to avoid this problem is maintaining up to date documentation; another is to avoid situations in which the sole source of knowledge regarding essential components of the software infrastructure is a single individual.

Appendix I

Charge to the STAR Internal Software and Computing Review Committee

The purpose of this review is to evaluate the requirements, the design and the ongoing (hardware and software) implementation plan for the STAR Software and Computing Project. In so doing, the anticipated physics program of STAR must be kept in mind. With only two years before STAR commences data acquisition, it is important to make sure that the software and computing approach (requirements, design and implementation plan) is correct and that there are sufficient resources in STAR to successfully carry out the required effort and thus satisfy the physics needs of STAR.

This review should be viewed as an opportunity for the Collaboration, as a whole, to gain an up-to-date and detailed view of all aspects of the STAR Software and Computing effort. This includes software infrastructure, track reconstruction software, simulations and analysis software, and the overall model for STAR computing at the RHIC Computing Facility, individual STAR institutions and the proposed STAR Computing Facility NERSC. In addition, the requirements of STAR in terms of CPU power and storage for offline computing for simulations, reconstruction and analysis should be evaluated.

Based on an open review, the committee should evaluate the software and computing design to determine whether the physics driven requirements will be met by the design and its expected implementation. The committee should consider the likely progression of physics-driven requirements from day 1 to year 1 and beyond and verify that the plan for developing the software and computing will result in the appropriate capabilities as a function of time; for pre-data taking simulations and testing of STAR, for turn-on, for one year later and for beyond when the full capabilities of RHIC and STAR are available. The evaluation should consider the physics requirements for both the nuclear physics program (heavy ion and related proton-ion and proton-proton studies) and the spin physics program and determine whether the software and computing fulfills STAR's requirements.

Appendix II

STAR Internal Software and Computing Review

October 20-22, 1997
3rd Floor Conference Room, RHIC Collider Center
Brookhaven National Laboratory

Agenda

Monday October 20 , 1997

Offline Software Requirements, Design, and Implementation

8:30 - 9:00 Executive Session

9:10 - 10:30 Overview, Requirements, Computing Model

9:10 - 9:40 Torre Wenaus
Overview of STAR computing.
Brief introduction to STAR
Organization, current personnel
Overall architecture GIF, Visio
Schedule and milestones PDF, MS Project
Computing resources overview

9:40 - 10:20 Peter Jacobs
Physics-driven computing requirements.
STAR computing requirements task force page

10:20 - 11:00 Torre Wenaus
Physics analysis. Torre Wenaus.
Brief comment on physics analysis organization and planning

11:00 - 12:30 Offline design, architecture, infrastructure

11:00 - 11:30 Torre Wenaus
STAR computing model
Data processing schematic GIF, Visio
Computing model schematic GIF, Visio
Working group report on the STAR computing model and plan

11:30 - 12:10 Craig Tull
The STAF Analysis Framework: Future of STAF
CHEP97 talk on STAF
STAF home

12:10 - 12:30 Torre Wenaus
Software Infrastructure

12:30 - 1:30 Working lunch

2:00 - 3:30 Simulation

2:00 - 2:10 Torre Wenaus
Introduction and Overview
2:10 - 2:40 Pavel Nevski
Status and plans for GSTAR
Including geometry, alignment/calibration database
2:40 - 2:50 Torre Wenaus
Status and plans for Geant4 in STAR
2:50 - 3:15 Bill Llope
Detector response simulations
3:15 - 3:30 Ron Longacre
Event generators. Ron Longacre

3:40 - 5:15 Reconstruction and Analysis

3:40 - 4:30 Lanny Ray
Reconstruction software
SVT contribution
4:30 - 5:00 Roy Bossingham
TPC cosmic ray test as an infrastructure and analysis testbed

5:30 - 6:30 Executive Session

<u>Tuesday, October 21, 1997</u>

Offline Computing Plan and Online Computing

8:30 - 9:00 Executive Session

9:10 - 10:30 STAR computing plan

9:10 - 9:40 STAR computing plan Doug Olson
RCF role and resources
Offline Computing for RHIC, 7/97, postscript, PDF
RHIC Computing Review 7/97 home page, reviewers' report
NERSC plans and status
Computing at the institutes
Working group report on the STAR computing model and plan

9:40 - 9:50 Torre Wenaus
STAR/RCF interface; RHIC common software

9:50 - 10:05 Torre Wenaus
Event store
Event store task force page

10:05 - 10:40 Doug Olson
Grand Challenge, data mining
Grand Challenge home page

11:00-12:30 General computing issues, planning, summary

11:00 - 11:10 Torre Wenaus

Visualization
Graphics page

11:10 - 11:30 Torre Wenaus
Infrastructure and development; session 2 leftovers
Code development tools and utilities
Code management and software releases
Quality assurance and control, bug management
Coding standards
Training
Documentation and communication

11:30 - 11:35 Torre Wenaus
Documentation and communication

11:35 - 11:45 Al Saulys
Code documentation

11:45 - 12:00 Torre Wenaus
Mock data challenges at STAR and RHIC

12:00 - 12:30 Torre Wenaus
Conclusion
Schedule, milestones, manpower: matching capability to requirements from day one through year N.

12:30 - 1:30 Working lunch

2:00 - 3:30 Online Computing

3:45 - 5:15 Online Computing

5:30 - 6:30 Committee meeting

Wed Oct 22 - Committee Meeting and Closeout

Morning: Questions from committee, committee meeting, report writing

Afternoon: Closeout session