## ADDITIONAL INFORMATION – PSCIC (Per request of C. Greer, 6/26/2007)

***(1) Site Visit Report*** *– Please provide a detailed response to each of the concerns and/or 'opportunities for improvement' identified in the report of the site visit team.*

*INTRODUCTION*

*"A revised budget will need to include the new educational initiatives and attention to the research needs of the social science community."*

**Response**: Revised social science and educational budgets and justifications, including Dolan DNALC, have been provided.

*VISION AND RATIONALE*

*"The site visit panel did not see any significant weaknesses in the vision and rationale for the project."*

*GRAND CHALLENGE IDENTIFICATION PROCESS*
*Opportunities for Improvement*

*The choices for "killer apps" that might be developed before grand challenge identification may benefit from community input so as to maintain the intent of the collaborative being "of, by and for" the community.*

**Response**: Killer apps will not be initiated unless the community 'buys in' to them in advance. Instead we will focus on developing a Discovery Environment prototype(s) in the first year. The possibility of killer apps will be raised at the initial Grand Challenge Conference to assess community needs and desires.

*Purpose and formatting of symposia beyond the first year may be best modified from the original plan so as to flexibly address feedback and the changing needs of the community and the collaborative.*

**Response**: Beyond the first year, both large and small conferences and symposia are dependent on community interest, input, and requests. We have altered our budget such that we now plan to hold only two large conferences, early in the first and third years.

*CYBERINFRASTRUCTURE*
*Opportunities for improvement*

*"The project team's response to questions raised in the initial reviews dramatically solidified the cyberinfrastructure plan. The clarifications and additional information provided in response fully satisfied the site visit team. The site visit team recognizes that there remain several difficult issues that pose challenges to any large scale project of this sort, and do require consideration. We recommend, therefore, that NSF work with iPlant to clarify their strategy for long-term archiving of irreplaceable datasets, and for quality control in the Discovery Environments."*

**Response**: Our thoughts on long-term archiving are contained in our answers to the four additional questions on the data management plan (see Section 3). With respect to quality control of the Discovery Environments, there are two issues. One is QC of the software infrastructure of the DEs. "Hardening" of the software is the primary responsibility of the cyberinfrastructure core's professional software engineering team. The team will use standard software development QA techniques including regression tests, unit tests, integration tests and acceptance tests to assure that the software does what it is supposed to across revisions and continues to be accepted and used by the target community.

The other aspect of QC is quality control on the contents of the datasets. Each data type will be associated with distinct quality control criteria. For example, DNA sequencing data will have QC criteria relating to trace length, percentage of Phred20 bases, and so forth. As part of the Grand Challenge Team DE requirements analysis and design phase, the IST team will identify the QC criteria that are appropriate for the datasets used by the DE in question. IST team members will then work with the infrastructure core to implement and apply these QC criteria uniformly across the datasets. Datasets that fail to meet the QC criteria will be flagged and referred to member(s) of the GCT for corrective action. In addition, provenance of the datasets will be tracked and made available to all users to make their own assessment regarding quality of the datasets.

*BUDGET AND TIMELINES*
*Concerns:*

*"The involvement of Barbara Heath and EMEC to do a formal independent evaluation of the iPC is important but still somewhat open ended in the sense that the measures described, while valuable, may not get at the most important issues, e.g. how are plant sciences being transformed by iPC? It may be inappropriate to expect EMEC to answer this. Thus the role of Susan Brown, a social scientist with expertise in studying information dissemination, will be very important. Her addition to the project is an asset, but it still wasn't clear how she would attempt to answer the key evaluation question, the degree and nature of the transformation (if any) of the plant sciences by iPC. This is admittedly hard as just what "transformation" would look like is vague. Nonetheless, it would be a recommendation to this team that this be studied, particularly with a view to identifying additional nonstandard, perhaps even innovative, measures that could indicate progress towards this ambitious but all important goal."*

**Response**: It is difficult to know what transformation will look like without having a more complete understanding of where the field is now. One thing that has been mentioned is that collaboration is not as well accepted in plant sciences as it is in other fields. The initial framework of the evaluation will follow a naturalistic design (Lincoln & Guba, 1985), which provides an opportunity for the question(s) related to the transformation of the plant sciences to naturally develop through purposeful data collection that is flexible enough to adjust based on analysis and preliminary findings. The evaluation and social and behavioral research teams will work closely to align efforts to attend to this critical goal. From the social and behavioral research perspective, one way to assess transformation is to see if (a) collaboration happens more and (b) if people's attitudes change.

For both the evaluation and research it is planned that during grand challenge meetings social scientists will interview some participants to get a better understanding of how people in the field think today. The interview questions will most likely be developed from the literature and a previously deployed survey sent to the general plant sciences population by the evaluation team. Given that, and what we hope the cyberinfrastructure will accomplish, we can develop more quantitative metrics for transformation. In addition, the social network analysis will give us a sense of how relationships change, thus providing additional insight into transformation.

*"An area of opportunity would be to explore innovative measures of social dynamics and to clarify the inter-community links that emerge from this effort. These will be critically important measures."*

**Response**: The primary issue associated with social dynamics will be similar to that mentioned for transformation. As the social dynamics in the field change, so should the social network. The recognition and tracking of inter-community links will be initiated through the evaluation of the IPC. Surveys will be developed to investigate both inter- and intra-community links. Preliminary analysis of the data will produce hypotheses that will be discussed with the social and behavioral research team providing them with an opportunity to design research projects related to community links and subsequently social dynamics. Measures associated with general attitudes and perceptions will be developed to understand how, if at all, social relationships are changing in the field.

*"It is strongly recommended that NSF annually evaluate expenses devoted to purchase of terabyte storage."*

**Response**: It has always been our intention to evaluate all infrastructure purchases, including storage, on an annual basis and to have them be driven by the needs of the community and the Grand Challenge teams. The proposed infrastructure development plan on page A3-2 of the proposal is merely representative and was used to determine reasonable costs.

*MANAGEMENT PLAN AND ORGANIZATIONAL STRUCTURE*
*Opportunities for Improvement*

*"The Board of Directors should meet with the University administration, at least annually, as an advocate for the project. The Directors of Cyberinfrastructure Development and Community Interactions should be full members of the Management Team."*

**Response**: The Board will meet annually with University administrators and the Directors will be full members of the Management Team.

*INTEGRATION ACROSS DISCIPLINES*
*Opportunity for improvement*

*"The iPC must absolutely continue to foster new and significant ties with international partners and industrial interests."*

**Response**: We agree this is critical. Steven Rounsley (Arizona) has lead responsibility for partnerships and will be assisted in international partnerships by Robert Martienssen (CSHL) and with industrial interests by Stephen Goff (Syngenta). As indicated during the site visit, dozens of potential partners, including international and industry, exist, and the main problem will be identifying the potentially most productive interactions. In addition, the PI and co-PIs will be proactive in this area and work with the Board of Directors to ensure appropriate international and industrial interactions.

*DIVERSITY OF PARTICIPATION*
*Opportunities for improvement*

*"The iPC will need to seek and continuously monitor the participation of individuals from very diverse and distant disciplines, such as physics, to ensure the richness and synergy of the collaborative."*

**Response**: We absolutely agree that participation by disciplines beyond plant biology and CISE will strengthen the collaborative and intend to structure the composition of the BoD to assist with these efforts. We will also take advantage of current interactions and connections existing between iPC faculty and researchers in the physical, mathematical and astronomical sciences. It is also a very important goal that there be broad participation of women, under-represented minorities, and faculty and students from the full spectrum of colleges and universities. See Diversity Plan below (#5).

*COMMUNITY AND PARTNERSHIPS*
*Opportunity for improvement*

*"Despite time constraints, it would be desirable to involve the community in the process of identifying the target application(s). (Killer apps)"*

**Response**: No effort towards 'killer apps' will made until and unless there is strong, diverse community input and support. Instead, during the first year we will focus on developing a Discovery Environment prototype(s).

EDUCATION AND TRAINING
*Opportunities for improvement*

*"Given the new participants (who substantially enhance the educational and training plan), we encourage a reconsideration of the budget allocated to this segment of the proposal. We also suggest that economically-disadvantaged populations be included as a diversity target."*

**Response**: A budget and budget justification for both Dolan DNALC and Purdue University have been provided, and the original higher education and K-12 budgets have also been increased. The EOT Plan below and our response to question #9 during the site visit provides many details of our approach to achieve broad participation. Our plan for broad representation, including women and under-represented minorities and high school, community colleges, 4-yr and research institutions will also target economically-disadvantaged populations. Two examples of this are 1) the K-12 school districts in Southern Arizona have 75-85% minority

students with an even higher percentage of students on the free lunch program, a measure of low income; 2) the 2-day workshops that run by DNALC targets the broad range of education institutions, including historically black and Hispanic colleges; all have a high percentage of students from economically-disadvantaged households.

***2) Revised Budget and Budget Justification***

To be provided separately.

***3) Data Management Plan***

*a. Page A4-1, Paragraph 1: "…we propose to develop a policy document that will outline principles, guidelines, and policies for access, protection and preservation". Please describe the process that will be used to develop this document and provide a timeline.*

We propose to develop a policy document that outlines principles, guidelines, and policies for access, protection and preservation of data. To accomplish this we will establish a working group composed of 20-25 volunteers from the community. The working group will meet at least 2-3 times in the first year during the grand challenge workshops to discuss and develop appropriate policies for data access, protection and preservation. The group will also review and incorporate existing recommendations from the Library of Congress and other communities that deal with public datasets such as space sciences, oceanography and others. The working group will deliver preliminary recommendations at the end of the second quarter and a final report at the end of the first year with their recommendations. The report will be widely disseminated in the community via the iPC. This working group will be chaired and coordinated by the Director of Cyberinfrastructure (DoC). The Board of Directors will approve the final policy document and the DoC will be responsible for implementing these policies. We recognize that data access and licensing polices may interact with each other, so the respective policy documents will clarify the nature of the interaction and resolve any conflicts that may arise. We will revisit the policies and mechanisms on a periodic basis via these working groups and propose revisions and refinements based on feedback and usage data gathered through the iPC.

*b. Paragraph 5: "… will create Discovery Environments (DE) and license them such that users cannot prevent others from using their improvements or additions to the DE." Please describe the proposed licensing strategy (e.g. Open Source [specify category], Creative Commons, etc.).*

The iPC will be producing substantial amounts of software of general utility to the community, creating integrated sets of data produced by third party groups, and participating in the production of novel data in the context of grand challenge team collaborations.

*Software.* To encourage the widest distribution of the software we produce, we will

distribute all novel software under the Academic Free License version 3.0 (http://www.opensource.org/licenses/afl-3.0.php), which allows software to be reused, modified and redistributed provided that attribution of the original work is maintained. Much of our software will depend on third-party components, however, and other licenses may apply to these dependencies. We will make every effort to use open source third-party components whenever technically feasible, and will provide the community with visible alerts in cases in which we have been unsuccessful in doing so.

*Integrated Data.* All datasets that we produce by a process of integration of third-party data will be distributed under the Creative Commons "by" 3.0 license ( http://creativecommons.org/licenses/by/3.0/) which allows the data to be copied, redistributed and adapted provided that attribution of the original dataset is maintained. In some cases, additional third-party restrictions will apply to the dataset. Whenever possible, we will avoid using restricted datasets due to the complexity of redistributing derivative works, but when unavoidable, we will alert the community to these restrictions.

*Novel Data.* Datasets derived by grand challenge teams will be released under terms of the "Fort Lauderdale data sharing agreement" (http://www.wellcome.ac.uk/assets/wtd003207.pdf) in which data producers release data publicly as soon as it is quality-control checked to assure high quality, and data users agree to cite the data producers and to respect the producers' rights to publish comprehensive analyses of the data in peer-reviewed papers. To assist the community in determining how to cite the data and which analyses the data producers are planning to publish, each grand challenge team will publish a "marker paper" describing the project, its citation policy, and its proposed analyses, at a stage when the initial project planning is mature, but before substantial work has been completed. Following publication of these analyses, the datasets will be made available without publication restrictions under the Creative Commons "by" 3.0 license (http://creativecommons.org/licenses/by/3.0/legalcode). Agreement with these data release terms will be a precondition for community member participation.

*Community software contributions.* All contributions in the form of software code, algorithm implementations that are contributed by the community for IPC projects, shall be accepted utilizing the agreement model and licensing templates adopted from the Apache Software Foundation (ASF). The "Contributor License Agreements (CLA) (http://www.apache.org/licenses/#clas), defines the terms under which intellectual property has been contributed to the collaborative and would be applicable to individual contributors, academic and industry groups and their employees. For software applications and documentation that are donated towards IPC projects by individuals or corporations will adhere to ASF "Software Grant Agreement (SGA) (http://www.apache.org/licenses/#grants).

*c. Page A4-2, Paragraph 4: "…it will be necessary to develop mechanisms and policies to decide what data to retain, remove, or archive and when to do so." Please describe the process that will be used to develop this policy.*

We will establish a working group composed of 20-25 volunteers from the community. This group will be tasked with developing policies and mechanisms for data and software retention, archiving and retirement. All of these issues fall under the general topic of information life cycle management (ILM). The working will have participants from diverse parts of the plant sciences and CISE community. As in part (a) above, the working group will deliver some preliminary recommendations by the end of the second quarter of year 1 and a final report by the end of year 1. We believe that there are several interesting research issues pertaining to automatic archiving and movement between storage devices. These research issues will be identified, clearly defined and investigated as part of the IST work in years 2-3. This working group will also be chaired and coordinated by the DoC. The Board of Directors will approve the final policy document and the DoC will be responsible for implementing these policies. We will revisit the policies and mechanisms on a periodic basis via these working groups and propose revisions and refinements based on feedback and usage data gathered through the iPC.

*d. Paragraph 5: "We will also develop a plan for disaster recovery and remote backups." Please describe plans for providing fail-over service to avoid loss of service and plans for disaster recovery and remote backup.*

The iPC will utilize the Living Disaster and Recovery Planning System (LDRPS) a software tool that is available to UA departments for developing and maintaining its customized Disaster Recovery Framework (DRF). Key measures utilized for this will include Recovery Time Objective (RTO) and Recovery Point Objective (RPO), which will dictate the acceptable amount of down time between the disaster and resumption of function (partial and complete) and allowable data roll-back. These measures will greatly depend on the values assigned to data along with business continuity needs defined by IPC subsections and Discovery Environments (DE), Grand Challenge (GC) teams, and community users. A well documented Service Level Agreement (SLA) will be made available to all constituents. Essential services such as authentication, information portal along with data and services defined as "mission critical" will be actively mirrored to off-site locations (details listed below).

The IPC disaster recovery plan (DRP) will build on top of existing plans defined by the participating institutions (UA and ASU). It will be derived in accordance with University Response Operation Plans and Guidelines and with Arizona State Government Information Technology Agency policies, standards and procedures, including the statewide Agency IT Disaster Recovery Policy (P051) and the statewide IT Disaster Recovery Standard (S901).

IPC equipment is to be housed at the Biotechnology Computing Facility (BCF) and the Center of Computing and Information Technology (CCIT) data centers at the University of Arizona and at the Fullton High Performance Computing Center at Arizona State University. (Note: ASU is located 100+ miles from UA). We will extend and leverage their infrastructures for fail-over services, described as follows:

a. ASU will act as off site storage and fail over service provider.
b. IPC will utilize technologies that facilitate addition of new off site (data and

service) providers to ensure a higher level of access.

c. For data with stringent RPO and RTO requirements, active off site replication will be utilized. This will include database cluster and disk-based mirroring. (BCF currently replicates all critical MySQL databases to a UA DR site.)
d. For services with stringent RPO and RTO requirements, services will be defined at configuration level to include a secondary site where date and services will be replicated.
e. BCF uses VMware "Infrastructure 3" virtualization technology, which allows provisioning of services and allocation and monitoring of resources. This includes the high availability (HA) component, which actively monitors servers and relocate services (CPU, Storage, Applications) to another server (location) in case of failure; it also has ability to migrate active servers (using Vmotion) and the ability to perform snapshots of server states and copy them to off-site location for quick recovery. Using virtualization technology eliminates the need to maintain identical hardware configurations to attain reliable high availability. IPC will inherit these best practices and tune them for its needs.
f. For data that will: a) not reside on shared file servers but on individual machines (desktop based analysis, remote users with limited bandwidth), b) require multiple versions archived, c) need "golden copy" stored in a non-active format, or d) reside on RAID/Mirrored storage but need copy for backup/DR purposes, the iPC will use the ALICE (Arizona Lifesciences Information Cataloging Environment) framework from BCF that comprises of tiered storage and data management rules enacted by IBM's Tivoli Storage Manager (TSM). The data will be migrated through the tiered storage down to tapes, which will be FedEx-ed to ASU (or other sites) on a monthly or quarterly rotation cycle. TSM also manages the server images and assists with "bare metal restore" in case of disaster.

***(4) Project plan*** *– Please provide a detailed plan, including tasks, primary responsibilities, deliverables/milestones, and timeline with completion dates for the startup phase and first year project activities. Provide an outline plan for project years 2-5. Identify risks to the project plan and describe plans for managing those risks. Describe the process to be used in managing the software development pipeline (including critical path and risk management). You may wish to consult with a professional project manager in developing your plans. Further, please indicate whether professional project management expertise will be available throughout the duration of the project and, if so, how this will be provided."*

We have consulted with a management expert, William VanLeuvan, MBA, Adjunct Lecturer, Department of Management Information Systems, Eller College of Business, University of Arizona, regarding project management and the management of risks. We have budgeted for a professional project management consultant throughout the project; the managing Directors of Cyberinfrastructure Development and Community Relations will obtain certification in professional project management; and members of the Management Team will obtain training in professional project management. We are evaluating Microsoft Projects 2007 software for use in managing the project; it provides very good intelligence tools

for building an executive dashboard.

Effective control of the project's *scope, schedule and budget* will be essential to the success of the iPC. The interrelationships among these three elements must be understood and managed at all levels of the project according to policies and guidelines established by the management team. Scope 'creep' is a major risk in this project; it will be critical that the management team clearly define scope for each project goal, including community goals, communicate scope definitions to staff and community, and regularly assess whether the scope is being modified in important, unplanned ways. Overall, it is essential for each significant goal of the project to organize tasks, manage level of effort, develop and monitor critical-path schedule, and compare actual cost to planned cost at key milestones. Progress toward timelines and relative to budgets will be assessed regularly; conclusions will be communicated effectively to staff and community participants; and adjustments will be made in consultation with appropriate staff and participants. The Directors of Cyberinfrastructure and Community Relations and the Collaborative will have key roles in assessing and managing these risks, in concert with the Management Team.

*Procurement risks* exist with respect to people, contracts, equipment, and services. Loss of *key personnel* is also a significant risk. The Management Team will develop a plan to identify all such risks and to develop means for addressing them. The Management Team will develop a plan to identify redundancies and backup plans in case a key project member is lost. Risks regarding *data and disasters* are addressed in the Infrastructure and IST plans below. Environmental scanning will be used to attempt to identify *external events*, including new technology developments, that could significantly affect the success and usefulness of the project. We would anticipate that major adjustments will have to be made at some point during the project because of new technologies. We will also perform a stakeholder analysis (in concert with EMEC) to assess how the project is perceived.

*Managing the software development pipeline* is a key concern because of the geographic relationship of CSHL and UA. The IST teams at CSHL and UA will use "agile" software development techniques to achieve rapid prototypes that have a high level of acceptance by the target users. Agile software development involves the following principles:

a. close, daily communication between software developers and users
b. creation of "user stories" early on to describe the scenarios under which the software will be used.
c. release of functional code early and often
d. tolerance of continuously changing requirements
e. close attention to technical excellence of code, including frequent code reviews, use of "code buddies," unit tests and regression tests
f. use of the simplest possible solutions to solve problems
g. flexibility to changing circumstances
h. reliance on a decentralized management model in which teams of creative individuals self-organize around practical challenges

The IST software development pipeline is geared towards development and prototyping of completely novel software. The prototypes, as they become functional, will be passed to the infrastructure core for hardening into production-quality software using a more traditional development methodology described later.

Due to the geographic relationship between the CSHL and UA IST groups, there is a risk that the two groups might perform as isolated projects rather than as an integrated team. To ensure that this does not happen, we will implement the same software development methodologies that have been successful in building coordinated teams of open source developers from disparate backgrounds and geographies. These methodologies include the use of a common source code repository, a feature request and bug tracking system, a shared mailing list, and a documentation WIKI. We will supplement these software features with regular "all hands" teleconferences to discuss strategy and progress on IST projects. In addition, co-PIs Stein and Ram will meet by telephone on a weekly basis to compare notes and discuss high level issues. We note that the Stein and Ware groups are already highly decentralized, with roughly a third of the software developers working from remote locations that are sometimes many time zones apart. The CSHL group is therefore well prepared to handle this highly integrated, but geographically decentralized project.

The Collaborative Director recognizes the essential necessity of effective, coordinated efforts between CSHL and UA and will pay especially close attention to managing this key relationship. Decisions about resource allocations between UA and CSHL (made with BoD oversight and approval) will depend strongly on the success of this relationship.

## A. Project Management Plan

### *Management, Oversight, and Communications Responsibilities:*

*Lead Manager*: Collaborative Director (Jorgensen)
*Collaborative Management Team*: Collaborative Director, Managing Director of Cyberinfrastructure (DoCI), Managing Director of Community Relations (DoCR), Stein (IST), Ram (IST), Andrews (Core Infrastructure), Ware (IST and Infrastructure), Chandler (EOT), Brown (Social), Rounsley (Synthesis Activities & Partnerships)
*Communications and Publicity*: iPC Communications Officer, Jorgensen, DoCR, DoCI

Regular meetings of the Collaborative Management Team will be held (weekly in the first year) to establish and implement team goals, policies, and lines of responsibility; to hire and manage personnel; to manage progress relative to timelines, budget, scope, and deliverables; to identify risks and develop plans to address them proactively; to report and respond to the Board of Directors; and to establish, maintain and extend external relationships and communicate effectively with the community.

***Objectives, Deliverables and Timeline***

*Pre-award startup phase*

1) Establish Nominating Committee (10/07)
2) Establish Board of Directors, chosen by Nominating Committee (12/07)
3) Staff hires: DoCI, DoCR, Communications Officer (initiate 10/07 and target completion for 1/08)
4) Design and begin implementing publicity plan to communicate and publicize iPC's goals, activities, and opportunities and to establish iPC leadership in promoting computational thinking in plant biology [(10/07)
5) Initiate preparation of materials to communicate and publicize iPC's goals, activities, and opportunities and to establish iPC leadership in promoting computational thinking in plant biology. First round of marketing and publicity materials ready 12/08, leading up to first GC conference in 4/08.
6) Create a unified iPC portal through which visitors will reach iPC information, community building tools, publicity materials, Discovery Environments, and other iPC tools as they are developed. Site will include several wikis for community input. [(10/07) see IS Team re this activity.]
7) Hire project management consultant, assess project management tools and software, etc. (1/08)

*Year 1*

1) Project management training: Directors and PI's (1/08)
2) Establish project management systems and procedures (1/08)
3) Establish project policies and practices (1$^{st}$ quarter)
4) Detail project scope, timelines and budgets (within project management system) for each team/area and for project as a whole; communicate these objectives clearly to teams and BoD; manage as necessary and appropriate (1$^{st}$ quarter and ongoing)
5) Evaluate risks and develop contingency plans (1$^{st}$ quarter)
6) Assess potential partnerships and establish as appropriate; initiate environmental scanning (1$^{st}$ quarter and ongoing)
7) Assess community (stakeholders) and communicate effectively: "build community" (1$^{st}$ quarter and ongoing)
8) Report and respond to BoD at least quarterly; BoD meetings 1/08; 4/08; 7/08; 10/08. Meeting in 4/08 to be held following Grand Challenge "kick-off" conference, either immediately after at CSHL or later in the month at UA.

*Years 2-5*

1) Review scope, timelines and budgets; communicate these clearly to teams and BoD; manage as necessary and appropriate (ongoing)
2) Evaluate risks and develop contingency plans (1$^{st}$ quarter)
3) Assess potential partnerships and establish as appropriate; initiate environmental scanning
4) Assess community (stakeholders) and communicate effectively: "build community"
5) Report and respond to BoD at least quarterly

## B. Core Infrastructure Team Project Plan

***Management and Oversight Responsibilities:***
*Lead Manager:* Director of Cyberinfrastructure (DoC)
*Core Infrastructure Management Team / Faculty Oversight:* Andrews (faculty lead), DoCI, Stanzione (ASU), Ware (CSHL), Hartman (UA), Merchant (UA)
*Communications*: iPC Communications Officer; coordinates with DoCID & DoCR

Regular meetings of Core Infrastructure Management Team will be held (at least biweekly in the first year) to coordinate activities at each site, including DE development, to respond to needs of the IST, to establish and implement goals, policies, and lines of responsibility, to manage progress to timelines, budget, and deliverables, to hire and manage personnel, to report and respond to the Collaborative Management Team, to establish, maintain and extend external relationships, and to discuss future developments. Members of the group will participate in iPC conferences and workshops to learn and understand the infrastructure needs of the community. The group will also interact with developers of other cyberinfrastructure projects to keep abreast of trends and best practices.

***Timeline and Deliverables:***

*Pre-award startup phase*

1) Post job announcements for 2 system administrators, 2 software developers, 1 programmer at UA and 1 system administrator at ASU

*Year 1, first quarter*

1) DoC assumes position; begins interviewing staff candidates
2) Describe detailed scope, timelines and budgets (within project management system)
3) Purchase laptops and workstations for core staff at all sites
4) Develop resource access/account policies
5) Start to form a resource allocation committee (with members from the outside community)

*Year 1, second quarter*

1) DoC and key staff participate in grand challenge kickoff meeting
2) Create specifications for storage and computation platforms.
3) Start building basic portal to allow users to access iPC facilities
4) Start work on first core projects (e.g., virtualization facilities to support checkpointing, migration, and archiving)
5) Post job notices for user support, DBA, 2 more software developers and another programmer at UA, and 1 programmer at ASU

*Year 1, third quarter*

1) Interview and hire additional staff
2) Start installing equipment (storage and cycles)

*Year 1, fourth quarter*

1) Gather input on needs for next year and start planning second year purchases and projects
2) Review staffing requirements and adjust plans for second year as appropriate
3) Deploy initial version of iPC portal (largely for information dissemination and basic collaboration tools at this point, as most modeling tools will be created in the DEs)
4) Make initial resource allocations

*Year 2*

1) Add additional infrastructure as determined by the needs of the community
2) Toward end of year hire additional staff as specified in the budget at both UA and ASU
3) Make computation and storage resources available to allocated users via iPC portal, and expose access mechanism to DE development teams.
4) Commence operation of iPC help desk facilities

*Years 3-5*

1) Operating at full staffing level.
2) The focus is on developing the hardware and software infrastructure to support the needs of Discovery Environments and Grand Challenge Teams.

***Risk Assessment – Cyberinfrastructure:***

1) The primary initial risk is a delay in being able to hire a qualified Director of Cyberinfrastructure. If this happens, we will appoint an Interim Director from among the personnel in somewhat similar positions at the University of Arizona.
2) Once the facility is operational, an ongoing risk is equipment failure or loss of data. We will mitigate these possibilities as specified in our disaster recovery and remote backup plan (see our answer to question d. on the Data Management Plan).
3) A potential future risk is that the infrastructure needs of the community overwhelm our resources and ability to provide adequate service. If this happens we will work with others, including NSF and external partners, to increase our funding and resource base.

## C. Integrated Solutions Team (IST) Project Plan

***Management and Oversight Responsibilities:***

*Lead Manager:* Director of Cyberinfrastructure Development(DoCID)

*IST Management Team / Faculty Oversight*: DoCI, Sudha Ram (UA), Lincoln Stein (CSHL), Doreen Ware (CSHL), Kobus Barnard (UA), Richard Snodgrass (UA), Rebecca Doerge (Purdue)

*Communications*: iPC Communications Officer; coordinates with DoCID & DoCR

Regular meetings of the IST will be held (at least biweekly in the first year) to respond to needs of the GC Teams and the Core Infrastructure Team, to establish and implement IST goals, policies, and lines of responsibility, to manage progress to timelines, budget, and deliverables, to hire and manage personnel, to coordinate DE development at each site, to report and respond to the Collaborative Management Team, to establish, maintain and extend external relationships, and to discuss future developments. Members of the group will participate in iPC conferences and workshops to learn and understand the informatic needs of the community. The group will interact closely with GC Teams and will monitor other projects to keep abreast of trends and best practices, new approaches, etc.

***Deliverables***:

Initially, the deliverables will consist of prototype software tools for data analysis, integration and management, algorithms for analysis and visualization, and workflow support, as well as the iPlant WIKI. After the first year, the deliverables will consist of prototype Discovery Environments designed and developed to meet community needs. The development of these prototype tools will be done collaboratively by UA and CSHL IST subteams.

***Timeline:***

*Pre-award startup phase*

1) Select iPlant WIKI and web site developers from existing programming staff
2) Establish iPlant web site front page & main sections (10/07)
3) Establish iPlant WIKI (10/07).
4) Initiate (11/07) and release (1/08) discussion documents on core datasets and services for community input.
5) Design (12/07) and release (1/08) web site and WIKI pages to support integration activities (e.g. support community proposals for grand challenge symposia)

*Year 1, first quarter*

1) Identify several initial core datasets to add to iPC, based on community WIKI discussions
2) Develop working groups for
   a. metadata descriptions
   b. data storage requirements
   c. workflow services
   d. algorithm adoption and development
3) Describe detailed scope, timelines and budgets (within project management system)
4) Schedule workshops for infrastructure standards & protocols.
5) Begin hiring new programming staff.
6) Develop processes and protocols for algorithm adoption and development

*Year 1, 2nd – 3rd quarters*

1) DoC and key staff participate in grand challenge kickoff meeting.

2) Hold workshops for developing metadata standards, policies for archiving, removing and retaining data and service level agreements, data access, protection, preservation.
3) Identify core algorithms for immediate adoption / development, based on community WIKI discussions and GC conference and workshops.
4) Work with newly formed programming teams to establish processes for the technical aspects of algorithm adoption / development.
5) Initial import of core datasets identified in first quarter.
6) Initiate prototype Discovery Environment development.

*Year 1, 4th quarter*

1) Begin working with Grand Challenge Teams (as soon as identified by BoD) to identify requirements.
2) Document requirements and begin design of first tranche of Discovery Environments.
3) Provide basic front ends for core algorithms identified in first two quarters.

*Years 2-5*

1) Gather requirements from community for Discovery Environments
2) Design and Prototype Discovery Environments working with GCT teams
3) Test and modify DEs based on community fedback
4) Document and pass on design requirements to Core Infrastructure team
5) Write research papers on CISE and PLS challenges in developing iPlant
6) Work with Evaluation and Social Sciences teams to collect data and analyze feedback and social networking in iPlant
7) Continue to innovate in developing new state of the art Data Management techniques, algorithms for analysis and visualization, workflow techniques.

***Risk Assessment – IST***

1) One risk is the lack of domain knowledge on the part of the IST staff. For example, if a grand challenge team involves extensive metabolomics, no current staff has the requisite domain knowledge. We will mitigate this risk by recruiting community domain experts to GCTs as collaborators. Flex funds will be used in this effort when appropriate.
2) Another risk is that core datasets and/or software identified by GCT will be unavailable, either because of technical reasons (e.g. they are stored in an inaccessible format) or because they are covered by non-negotiatable intellectual property restrictions. We will avoid this situation during the GCT feasability study phase, when a key criterion for feasability will be the unencumbered availability of key data and software needed to support the GCT project.
3) A third risk involves the use of ontologies to integrate information from diverse sources, a service that is a core feature of DEs. While it is to be hoped that communities will spontaneously reach a consensus on standardized ontologies, this is not always the case and the IST may be forced to make a choice between one ontology and another. To reduce the risk that we choose wrongly, we will work closely with collaborators at the

National Center for Biomedical Ontologies to evaluate competing ontologies using generally-accepted criteria for quality and completeness when choosing among competing ontologies.

4) Because of the decentralized nature of the IST effort, there is a risk that a DE software component designed by one team or team member will not interoperate with another component designed by a different team or team member. To minimize this risk, we will make extensive use of unit tests during prototyping, as well as a philosophy of releasing working code early and often. This will allow us to monitor whether each piece of software is doing what it is advertised to do and to test directly whether interoperation is working in practice as well as in principle.
5) Finally, there is a risk that the software systems designed by the IST do not win acceptance in the community because of a failure of imagination to anticipate the community's needs. We will seek to avoid this situation by engaging in extensive consultation with community members at first during grand challenge meetings and later during GCT formation. In addition, by releasing prototype DEs early and often, we will have an opportunity to evaluate community uptake at an early phase of software development by tracking usage of the software.

## D. Synthesis Activities Team Project Plan (Grand Challenge ID Processes)

### *Management, Oversight, and Communications Responsibilities:*

*Manager*: Director of Community Relations (DoCR)
*Synthesis Activities Management Team (SAMT) / Faculty Oversight*: Jorgensen (lead); DoCR, Rounsley, Wing, Martienssen, Stein, Doerge, Sanderson, Enquist
*Communications*: iPC Communications Officer; reports to DoCR

Regular meetings of the SAMT will be held (at least biweekly in the first year) to establish and implement goals, policies, and lines of responsibility; to hire and manage personnel; to manage progress to timelines, budget, and deliverables; to report and respond to the Collaborative Management Team; to assist the BoD with evaluation of proposals; to organize conferences and workshops and assist Grand Challenge Teams in preparation of proposals, plans, interactions with the IST, and visits to the iPC centers at UA and CSHL; and to establish, maintain and extend external relationships with the community. Members of the group will organize and participate actively in iPC conferences and workshops to learn and understand the cyberinfrastructure and computational needs of the community. The group will interact closely with GC Teams and will monitor other cyberinfrastructure and computational projects to keep abreast of trends and best practices, new approaches, etc., and to identify collaborations and synergies with potential partners.

***Timeline and Deliverables:***

*Pre-award startup phase*

1) Solicit white papers from the community for the 'kick-off' Grand Challenge Conference
2) Recruit Conference Organizer and staff
3) Prepare "introductory" Marker Papers and submit to major journals

*Year 1, 1st quarter*

1) Hire Conference Organizer and staff
2) Assist BoD in evaluation and external review of white papers
3) Make invitations to session organizers and invited participants on behalf of the BoD.
4) Make invitations to community for attendance or participation by webcast.
5) Plan and hold Grand Challenge Conference (April 7-11 at Cold Spring Harbor Lab; Grace Auditorium)
6) Prepare Marker Papers and submit to major journals as appropriate throughout the year

*Year 1, 2nd quarter*

1) Work with community conference session leaders to ensure production of comprehensive summary of GC Conference and display results at project web site, along with archived webcast of Conference proceedings
2) Solicit white papers/proposals for focused Grand Challenge Symposia/Workshops and work closely with proposers as needed/requested to ensure complete, comprehensive proposals that meet project expectations
3) Assist BoD in obtaining evaluations, prioritizing proposals; make invitations on behalf of BoD

*Year 1, 3rd and 4th quarters*

1) Hold at least 6 workshops during first year
2) Solicit Grand Challenge Team proposals for Discovery Environment development and work with proposers as needed and requested
3) Facilitate interactions with IST to develop DEs and host GC Team visits to iPC in BIO5.

*Years 2-5*

1) Continue organizing focused Grand Challenge Symposia as needed (3-6 per year)
2) Hold second Grand Challenge Conference in year 3
3) Facilitate interactions with IST to develop DEs and host GC Team visits to iPC in BIO5.
4) Publish Marker Papers with GC Teams and IST to describe DEs

***Risks Assessment - Synthesis Activities***

The principal risk in this area is that of community perception and 'buy-in'. It is critical that the community understand the nature of the project and the

opportunities offered it and participate actively and constructively. This will be addressed through a vigorous publicity and community relations effort immediately upon approval of the project (see 7, Publicity Plan). A lack of useful white papers would be an immediate concern and would require vigorous effort to obtain community contributions, and failing that, we would shift to a new approach that is necessarily more project and BoD driven to ensure broad representation at the kick-off conference for disciplines, approaches, ideas, etc. In addition, close interaction with the social and evaluation managers (Brown and Heath) will be necessary in the first year in order to understand and respond to community perceptions (see sections F and G). Brown will be a member of the Management Team and Heath will participate in meetings of the Management Team to ensure close, informed management of this risk area.

## E. Education Outreach & Training (EOT) Plan

***Management, Oversight, and Communications Responsibilities:***

*Manager*: Managing Director for Community and Education

*Faculty Oversight*: Chandler (lead): Napoli (K-12), Westbrook (higher ed), Stapleton (iPATs), Micklos (DNALC); Doerge (STATCOM)

*Communications and Publicity*: iPC Communications Officer; reports to DoCR

Regular meetings of this group (known as EOT Management Team) will occur monthly via video conferencing to establish and implement goals, policies, and lines of responsibility, to hire and manage personnel, to report and respond to EOT Advisory Board, and to establish, maintain and extend external relationships. Collaborative Director will be present at all such meetings.

***Pre-award Objectives/Deliverables:***

1) Establish EOT Advisory Board with BoD approval by 12/07.
2) Initiate new staff hires at CSHL and UA [10/07]
3) Initiate search and hire iPC educational coordinator at DNALC and four positions in support of EOT at UA: Education and Outreach Coordinator (50% grant funds/50% match), K-12 Educator (50% match), Learning Technologies Center developer (UA match), and UA iPAT coordinator (50% grant funds). [1/08]
4) Design and begin implementing publicity plan to communicate and publicize iPC's EOT goals, activities, and opportunities and to establish iPC leadership in promoting computational thinking in plant biology to the K-20 education communities. [(10/07) Initiate preparation of materials to communicate and publicize iPC's EOT goals, activities, and opportunities and to establish iPC leadership in promoting computational thinking in plant biology to the K-20 education communities. First round of marketing and publicity materials ready 3/08, in time for GC conference in 4/08.
5) Create a unified iPC portal through which visitors will reach publicity and EOT materials, community building tools, Discovery Environments, and iPC tools as they are developed. Site will include an EOT wiki for community input. [(11/07) Design and establish EOT interface within the iPC website, including an EOT WIKI for community input. Will work with IS Team on this activity; IS

Team will implement necessary website developments.

***Year 1 Objectives/Deliverables:***

1) Obtain EOT Advisory Board input on all iPC EOT initiatives. [2/08: Hold inaugural EOT Advisory Board meeting to present goals and plans for the educational, training and outreach components of iPC; 3/08 Adjust goals, based on EOT input. Lead Chandler and Jorgensen; key participants Micklos, Napoli, Stapleton, Westbrook, Doerge]
2) Communicate and publicize iPC's EOT goals, activities, and opportunities and to establish iPC leadership in promoting computational thinking in plant biology to the K-20 education communities.
    a. Attend 2008 conferences and symposia to obtain community input, promote iPC opportunities and organize EOT sessions at all conferences where possible [01-12/08 each team member will attend at least one conference Chandler, Westbrook, Stapleton, UA iPAT Coordinator, Napoli, Micklos, Doerge]
    b. Identify 2009 conferences to propose panels and sessions on CT in biology & iPC [05/08 Contact 10 appropriate conferences with CT suggestions for program; Westbrook, Stapleton, Napoli, Micklos, Doerge]
    c. Develop Internet multimedia to publicize the project and to involve students and teachers in the Grand Challenge development process. [1-6/08 Develop a meta-tag and content management system for educational content. 8/08 Test the alpha Internet site and release beta version. Continue usability and audience testing throughout the year. 5/08-12/08 Produce online videos and audio/video podcasts to introduce researchers and students to Grand Challenges in Plant Biology. 3-12/08 Develop a prototype education interface to plant genome analysis tools. Cooperate in the development and integration of K-12 and higher education content generated by EOT team members. Lead Micklos, Stein; key participants Napoli, Stapleton, Westbrook]
    d. Use Grand Challenge symposia and workshops as sources for video interviews, podcasts, and animated features. [2/08 Conduct video interviews with iPC managers and NSF program officers on project vision and goals (in conjunction with initial advisory board meeting). 3/08 Produce and publish online videos and audio/video podcasts on project vision and goals. 4/08 Conduct video interviews of speakers and attendees of Grand Challenge symposium and workshops. 5/08-12/08 Produce narrated videos that explain the conceptual background and historical development of each Grand Challenge. Lead Micklos]
    e. Document the development of a cyber- and community-based research paradigm. [7/08 and 9/08 conduct video interviews of speakers and attendees of educational workshops. 3-12/08 Produce online videos and audio/video podcasts of Educational Challenges in Plant Biology. Lead Micklos]
3) Initiate UBRP program in CT with 4 summer undergraduates [2/08 identify potential faculty mentors for students. 3/08 advertise program broadly to

undergraduates and faculty. 4/08 choose applicants. 5/08 initiate summer program. Design and give seminars on CT in Biology to UBRP and other interested students throughout year. Lead: Westbrook and Bender (UBRP coordinator)]

4) Carry out first summer internship program for 6 K-12 teachers and 3 high school students. [2/08 launch publicity and application materials; 3-5/08 develop program, identify faculty mentors; 4/08 review applications and select participants; 6-7/08 summer internships. Lead: Napoli]
5) Hold K-12 and higher education workshop in CT in Biology [1/08 confirm venue; 2/08 establish programs and confirm invited speakers; 3/08 initiate publicizing workshops widely; 8/08 confirm participants; 10/08 hold K-12 and Higher Ed workshop. Lead: Napoli and Westbrook.]
6) Initiate STATCOM at UA [1/08 Doerge (Purdue) and Piegorsch (UA) will initiate discussions with statistics students and faculty at UA. Interested UA students will be introduced via email to Purdue STATCOM student administrators who will serve as mentors and provide necessary materials for establishing STATCOM (e.g., CDs, written material, timelines). 02/08-07/08, STATCOM representatives from Purdue will continue to interact via email and phone with UA students in a mentoring role. A student representative from Purdue STATCOM will travel with Doerge to UA during year 1 (07/08) to help flesh out, and direct the interests of UA STATCOM. Also during year 1, Doerge will initiate discussions with a network of statisticians, who are currently working in plant science within NSF Plant Genome funded institutions. They will be made aware of the iPC activities and encouraged to participate in iPC conferences. The goal is to bring in statisticans (paid from flex funds) on iPC grand challenge projects. The long-term vision is to then (via Purdue) grow STATCOM in their institutions.]
7) Establish policies for marketing and selection of iPATs and initiate development and testing of iPAT training materials. [3/08 develop criteria for iPAT selection, disseminate publicity and application materials beginning with Grand Challenge symposium in 4/08. 12/08 develop training materials for iPAT consulting visits and system for tracking applications, visits and completion of each-one-teach-one responsibilities. Lead: Stapleton and UA iPAT coordinator]

***Year 2-5 Objectives (2009-2012):***

1) EOT Advisory Board meetings. [Hold annual Advisory Board meeting to summarize progress, receive input and adjust plans accordingly for the educational, training and outreach components of iPC. Lead Chandler and Jorgensen; key participants Micklos, Napoli, Stapleton, Westbrook, Doerge]
2) Evaluate progress in meeting previous year's objectives and refine all programs based on evaluation results. [Formal evaluation EMEC; Lead: Chandler and Jorgensen; input Micklos, Napoli, Stapleton, Westbrook.]
3) Continue EOT publicity, material distribution and community building activities
    a. Continue development and testing of the unified iPC portal. [Lead Micklos; key participants Napoli, Stapleton, Westbrook]
    b. Make iPC technologies and tools available to a broad audience of

biology faculty and students. [Lead Micklos; key participants Napoli, Stapleton, Westbrook]

c. Plan and conduct one higher educational symposia and one K-12 symposium each year. Lead: Napoli and Westbrook; key participants Micklos, Stapleton]
d. Continue to document the Grand Challenge development process, and to use symposia and workshops as sources of Internet content and other marketing materials. [Lead Micklos; key participants Napoli, Stapleton, Westbrook]
e. Integrate K-12 and higher education content including K-12 and Higher Education teaching modules for web dissemination [Lead Napoli (K-12), Micklos and Westbrook; key participant Stapleton]
f. Conduct workshops to introduce the iPC portal and educational curricula to high school and college faculty, including substantial proportions of Hispanic and Black faculty and those teaching in areas with high numbers of economically disadvantaged students. [Lead: Micklos]

4) Initiate and integrate iPAT teams within Discovery Environments and Grand Challenge teams: 4 in year 2; 6 in year 3 and 8 teams each in years 4-5 [Lead: Stapleton; each year consulting visits and training materials will be refined based on participant feedback and formal evaluation.]
5) Expand UBRP opportunities to national pools of students and integrate participants into iPC research activities [Lead: Westbrook and Bender. Goal is 7 students for year 3 and 10 students per year for years 4-5.]
6) Initiate international undergraduate student experiences through BRAVO! [Lead: Westbrook and Bender. Goal is 1student for year 2 and 3 students per year for years 3-5.]
7) Continue STATCOM program at UA and Purdue; expand STATCOM program to other universities [Doerge, Piegorsch and UA and Purdue STATCOM students, as described in year 1 above; goal is one new program in year 2, and two new programs per year in year 3-5.]
8) Expand K-12 teacher and high school student summer internships to include national pools of students and integrate participants into iPC research activities [Lead: Napoli; goal is 9 teachers and 6 students for summer internships each year.]

***Risk Assessment - EOT:***

A primary risk for first year could be delays in hiring one or more of the 5 staff positions (4 at UA and 1 at CSHL). If this happens we will adjust the responsibilities of the staff that are hired and also allocate time from existing staff within BIO5 at UA or DNALC to support the leadership team in moving the projects forward. Another risk is the loss of one or more faculty leaders, i.e. Napoli for K-12, Stapleton for iPATs, etc. One factor that mitigates this risk is the backup expertise within the leadership team, which is sufficient to keep the programs running on an interim basis while a permanent leader is identified. For example, the associate director at DNALC can fill the role of Micklos if necessary and vice versa;

the UA iPAT coordinator and Stapleton and Westbrook and Napoli are backups for each other.

## F. Social and Behavioral Research Plan

***Management, Oversight, and Communications Responsibilities:***
*Manager*: Director of Community Relations (DoCR)
*Faculty Responsibilities/Oversight*: Brown, Jorgensen

Susan Brown will participate in weekly Management Team meetings in year 1 to maintain current understanding of project development which will be reflected in the development of the evaluation plan. Interaction with Management Team in year 1 to provide input to the project development process. Additional planning meetings with the Evaluation Research Team will be scheduled to align data collection efforts.

***Deliverables***:

The key deliverable will be data – data to provide input to the development process, as well as to answer social science research questions.

***Pre-award Objectives/Deliverables (Fall 2007)***

1) Collect baseline social network data
    a. Distribute social network survey to plant science community
    b. Obtain publication and grant data for bibliometric portion
2) Collect data regarding innovativeness of plant science community members
3) Conduct preliminary usability analysis on user interface and provide input to the development team

***Year 1 Objectives/Deliverables (2008)***

1) Work with IST, Synthesis Team, and grand challenge groups to develop broader understanding of plant science community's needs and uses of iPlant.
2) Provide results of innovativeness survey to management team for identification of key users
3) Prepare survey for users of iPC web site, Wiki, and other tools
4) Conduct large scale survey for initial users of the iPC.
    a. Results of usability analysis will be fed back to design team
    b. Attitudes toward the iPC will be used to promote the iPC to future users and direct project management.
5) Hold two workshops for social science researchers to generate research questions and develop data collection needs.
6) Work with development team to build data collection into the system

***Years 2-5 (2009-2012)***

1) Monitor use of the iPC web site, Wiki, and Discovery Environments
2) Continue surveying users based on data needs identified by social scientists
3) Continue providing feedback to development team for system enhancements

4) Two additional large scale social network surveys and bibliometric analyses
5) Two additional workshops with social scientists
    a. to generate additional questions
    b. to identify additional data needs
    c. to conduct preliminary data analysis of iPC data

***Risk Assessment – Social and Evaluation:***

The biggest risk is the lack of data caused by a lack of responsiveness to the surveys – paper or online. Should this happen, several tools can be employed to address this problem:

- For online surveys, we could restrict access to the iPC web site based on survey participation (i.e., it is the price to join), but the downside of this is that it could reduce participation with the system, so this will be done optionally for basic entry and only higher level activities will require participation.
- We could offer incentives to encourage participation (e.g., Amazon gift certificates, etc.).
- We can rely on the social influence of the plant science community to encourage people to participate in the surveys.

## G. Evaluation Plan

***Management, Oversight, and Communication Responsibilities:***
Manager: Barbara P. Heath
Oversight: Board of Directors (BoD)
Communication: Attendance at the weekly Management Team meetings in year to maintain current understanding of project development which will be reflected in the development of the evaluation plan. Additional planning meetings with the Social and Behavioral Research Team will be scheduled to align data collection efforts.

***Pre-award Objectives/Deliverables***

1. Track pre-project development and progress (personnel hiring, board development, intra/inter-group communication, website/WIKI development, publicity efforts, etc.)
2. Begin related literature reviews
3. Collaborate with social networking team to align efforts and complete IRB requirements
4. Complete preliminary evaluation plan and submit to management team
5. Draft initial surveys, metrics
6. Complete any interim formative reports for project team
7. Search, interview, and hire staff for project start date

***Year 1 Objectives/Deliverables***

1. Attend project meetings, kick-off conference, and other major project activities

2. Create additional data collection methods to based on observational data
3. Collect data (documents, surveys, meeting attendance) related to each collaborative component, such as:
    a. Collect meeting minutes and planning documents from various teams
    b. Interview core staff post-kick-off meeting
    c. Create initial survey email lists
    d. Survey community at large (in collaboration with social science team)
    e. Survey participant community (after conferences/workshops)
    f. Track K-12 internship efforts, survey participants
    g. Gather usability data from EOT and social networking teams
    h. Follow publicity efforts
    i. Track policy/protocol development (various)
4. Analyze data, produce reports for project teams after major collection activities
5. Review data collection methods and revise as necessary
6. Produce and submit annual formative report

***Year 2-5 Objectives/Deliverables***

1. Attend project meetings and major project activities
2. Collect data (documents, surveys, meeting attendance) related to each collaborative component
3. Analyze data, produce reports for project teams after major collection activities
4. Review project development, add or subtract data collection as necessary
5. Produce and submit annual formative report

***Year 5+ Objectives/Deliverables***

- Collate data from years 1-5
- Analyze data as complete set to inform summative responses
- Produce and submit summative report and all data to management team

***Risk Assessment***

See Risk Assessment in Social and Behavioral Research Plan

***(5) Diversity Plan*** *– Your proposal includes important goals in responding to the requirement of the solicitation to "increase the participation of women and individuals from groups underrepresented in science and engineering at all levels and in all Collaborative activities." Please describe who would have primary responsibility for aggressively pursuing those goals and where this individual(s) would be placed within the management plan and organizational structure.*

Professor Vicki Chandler will be the project's Diversity Officer, taking primary responsibility for aggressively pursuing the goals of increasing the participation of women and individuals from groups under-represented in science and engineering in all Collaborative activities at all levels. Professor Chandler is a co-PI on the iPC proposal, is a member of the Management Team, and is responsible for leading faculty oversight of all iPC education, outreach and training activities. She reports to the Collaborative Director, Richard Jorgensen and will work closely with the

external evaluator, Dr. Barbara Heath, who will carry out collection and analysis of data that will inform on the successes and shortcomings of each component's diversity goals. Recruitment goals and metrics within each aspect of iPC will be set in consultation with our advisory board. Annually, recruitment of women and other under-represented groups and participation by people from all types of institutions and from economically disadvantaged backgrounds will be evaluated within each component of this project. Data will be collected using multiple methods and in collaboration with social scientists involved with the Collaborative. When necessary, changes to the project plan will be recommended and implemented based on the data collected. More details of our approach are in the response to question #9 from the site visit.

***(6) Evaluation Oversight and Response Plan*** *– Your proposal and site visit presentation describe the role of a contractor (EMEC) in ongoing evaluations of the project, including collecting usage statistics and conducting interviews and/or surveys. Please describe plans for contract management (including who would be responsible for contract oversight). State who would own the materials delivered under the contract and whether access will be open to the community or restricted (if the latter, indicate the nature of planned restrictions). State whether IRB review(s) is/are needed for the planned evaluation procedures and, if so, when approval(s) is/are expected. Describe how the evaluation materials would be used by the project team to improve the Collaborative.*

The Managing Director of Community Relations will directly manage the EMEC contract, and will report on this to the Collaborative Director and the Board of Directors. The Collaborative Director and the Board of Directors will have oversight responsibility for the contract. Ultimately, it is the CD's responsibility to ensure that the contract conditions and expectations are met in full. The University of Arizona will own the materials delivered under the contract and access will be open to the community with the only restrictions being regulatory. IRB review will be required and will be initiated by UA in August, 2007, and approval is anticipated in October, 2007. The Collaborative Management Team will respond to information in the evaluations that could be used to improve the Collaborative by developing and implementing plans for revision/modification of Collaborative objectives, procedures, practices and policies, and will report on this to the BoD.

***(7) Publicity Plan*** *– Please describe a detailed plan that could be used in publicizing a launch should an award be made. Indicate how that plan could ensure that all the relevant communities are informed. Describe the process and timelines for 'marker papers' as proposed in the site visit. Note that NSF expects to work with any awardee in preparing a joint press release in the event of an award.*

## A. Media Relations

1) **Joint press release.** Upon approval, we will work with NSF to develop a joint press release, with UA acting as lead institution, coordinating with the subawardee institutions. The release will be distributed via international wire services including Newswise and EurekAlert in addition to a national Vocus media database service, and University media lists. Personal pitches will be

made to prominent science and higher education reporters. A plan to have 24 hour availability of spokespersons to respond to media inquiries will be defined. Spokespersons will receive media training prior to the distribution of the press release. Satellite interviews with non-local TV stations and teleconferences with non-local print publications will be made available.

2) An **electronic media kit** will be developed and placed on the iPC web site; it will provide comprehensive information about the program that reporters can use for background.
3) A **press conference** will be held to coincide with the announcement. Ideally, representatives from NSF and the iPC team will participate. The conference will be video streamed on the Media Relations section of the iPC web site.

**B. Publicizing and Communicating iPC goals to the Community**

We recognize that timely, thorough, broad and inclusive publicity about iPC is crucial for early community buy-in and we will work with professional marketing and communications staff at UA and CSHL to develop a thorough plan for the initial launch (and ongoing) publicity. The dynamic project web site will be a key portal for distribution of communication materials and the project WIKI will play a key role in facilitating and promoting community dialog about the iPC. A database will be maintained of target audience members in two sections: a) those requesting to be on list for regular e-mail updates, notices, and possibly an e-newsletter and b) those we identify as potentially interested community members whom we would like to notify about the iPC. Information requests, including those received via the WIKI, will be addressed and provided to all via the project web site. Information will also be provided to professional societies, which will be encouraged to provide links at their web sites, to diverse research institutions, targeting specific Departments, Colleges and Programs, and to major databases and web-based communities in plant, computer and information sciences, which will be encouraged to provide links. Key materials that we will produce are as follows:

1) A **webcast presentation** will be made for the community by the PIs and appropriate key personnel to explain the nature and opportunities afforded by the iPC and to answer questions about the processes and avenues for participation, responsibilities of the iPC and the community, respectively, etc. The webcast will be available at the project website, including podcast format, along with the 'long format' project description. In addition, webcast question and answer sessions will be scheduled as needed.
2) A **'long format' project description** (approximately 10 pages, plus online supplementary material) which, together with the webcast will a critical 'road map' for community members to understand the nature of the iPC, the nature of the opportunities for participation, the processes through which the community can provide input, debate priorities, and make decisions, the nature of services that the iPC expects to be able to provide, and the responsibilities of the community and the Grand Challenge Teams.
3) **Other types of materials** we envision distributing include flyers and brochures for mailing and distribution at scientific meetings, e-mailings to broad memberships with permission and assistance from professional

societies, the iPC website, linked from NSF, professional society and community websites, a WIKI for community input and dialog, and online videos and podcasts targeted for diverse communities (produced by Dolan DNALC; see EOT #2c,d,e in year 1).

**Seminars** at research institutions and **presentations** at scientific conferences, afford opportunities for more detailed and personal communication of collaborative objectives and opportunities to the community. A "request form" will be placed on the iPC web site so that interested groups can request a speaker for their event/seminar. A core group of speakers will be established, trained and provided with collateral materials supporting the program and with an electronic presentation that is audience appropriate. Outreach efforts to solicit speaking opportunities will also occur. Examples of the opportunities we intend to exploit include:

1) The Collaborative Director (CD) will offer to make presentations at research institutions and at scientific conferences. A presentation on the iPC and grand challenges in plant genetics and epigenetics will be given at the annual meeting of Agronomy and Crop Science Societies of America on November 7, 2007 in New Orleans. The possibility of a workshop will be discussed with the organizers.
2) The CD also plans to attend the Genomic Informatics conference at Cold Spring Harbor, Nov. 1-5, 2007, and will submit an abstract for oral presentation as well as offer to hold an informal workshop to be co-presented with Lincoln Stein and Doreen Ware.
3) Rod Wing, an organizer of the annual Plant and Animal Genome conference in San Diego, January 12-16, 2008, will schedule a workshop on the iPlant Collaborative and Grand Challenges in Plant Science, repeated annually.
4) Rod Wing, organizer of the annual International Rice Functional Genomics Conference in Tsukuba, Japan, October 14-17, 2007, is organizing a day-long workshop on Grand Challenges in Plant Biology on October 18, immediately following the conference.
5) Annual meetings of professional societies in biology, such as ASPB, BSA, and ASA-CSSA-SSSA, and computational and information sciences, such as IEEE, ACM and ACS, will be offered presentations or workshops on the iPC and the opportunities that exist for the community. All project faculty may contribute in this way at appropriate conferences.
6) The PIs will exploit standard seminar invitations to universities as a means to present the nature and opportunities of the iPC. For instance, a seminar invitation to the CD at the Danforth Center in St. Louis will be scheduled for January, 2008, with side trips to other institutions, such as University of Missouri, Columbia. Another such opportunity is UCSD/Salk in conjunction with the PAG conference in San Diego.

We will also use a system of "**marker papers**" to engage the community and keep them apprised of iPC activities. The marker paper concept was first formulated by the genome sequencing community at the Wellcome Trust-sponsored Fort Lauderdale meeting (http://www.wellcome.ac.uk/assets/wtd003207.pdf) in 2003. Originally called a "Project Description," a marker paper is published in a high

profile journal suitable for the target community at the beginning of a project. The contents of a marker paper includes: 1) the scientific rationale of the project and its major strategies; 2) the expected deliverables and timeline from the project, including community access to information resources generated by the project; 3) the nature of the major analyses that the project will perform. The purpose of the marker paper is threefold. 1) It encourages community involvement in the project; 2) it provides a citation for the community to use when using unpublished data sets produced by the project; 3) it declares the intention of the project to perform a series of large-scale analyses by a declared date, thereby avoiding duplication of effort.

We will publish two types of marker paper. The first marker paper will announce the creation of the iPlant Collaborative itself. It will describe the structure of the collaborative, explain the open nature of access to iPC resources, explain how strategic decisions will be made, and invite community members to get involved in this process. This paper will be published in one or more of the following high-profile journals:

a. Nature
b. Science
c. Plant Physiology
d. CACM (Communications of the ACM)
e. IEEE Computer
f. CAIS (Communications of the AIS)
g. ISR (Information Systems Research)

In order to reach biologists and computer scientists/engineers equally, we will create two initial marker papers; one describing the project from the point of view of grand challenges in plant sciences and published in a journal that has greater reach to biologists (e.g., *Nature*), and the other describing the project from the point of view of grand challenges in computer science and published in a journal with greater reach to the CISE community (e.g., *CACM).*

The second type of marker paper will appear soon after a Grand Challenge Team is constituted, one per Grand Challenge Team. It will describe the grand challenge problem, the strategy to be used to address the problem, the resources to be generated and the expected timeline, and the analyses that the GCT intends to perform. These papers will appear in the journal most suitable for the Grand Challenge problem chosen.