THE HARVARD INSTITUTE FOR INNOVATIVE COMPUTING

NOVEMBER 22, 2004

Principal Authors: Timothy Clark (HMS), Alyssa Goodman (FAS) & Christopher Stubbs (FAS)

Principal Editors: Felice Frankel (MIT), John Huth (FAS), Hanspeter Pfister (Mitsubishi Electric Research Laboratory), Joy Sircar (DEAS)[1]

TABLE OF CONTENTS

MOTIVATION 1

EXPECTED BENEFITS OF THE IIC 2

AREAS OF IIC ENDEAVOR 3

IIC STAFFING AND ORGANIZATIONAL CULTURE 7

PLANNED STRUCTURE & EVOLUTION 8

IIC LEADERSHIP: START-UP TASKS AND APPROACH 10

APPENDIX A: A REPRESENTATIVE SAMPLING OF IIC PROJECTS 11

APPENDIX B: ONE-PAGE DESCRIPTIONS OF SAMPLE IIC PROJECTS 12

PARSING, DISPLAYING, AND SERVING THE COMPLETE DATABASE ON THE WEB: A STEPPING STONE TO THE VIRTUAL OBSERVATORY 14

APPENDIX C: SAMPLE RESEARCH INSTITUTES AND INITIATIVES AT HARVARD WHOSE MEMBERS WOULD COLLABORATE WITH THE IIC 17

APPENDIX D: OUTSIDE EXPERTS WHO HAVE AGREED, IN PRINCIPLE, TO SERVE ON THE IIC'S EXTERNAL ADVISORY COMMITTEE 18

APPENDIX E: FULL LIST OF LOCAL RESEARCHERS PARTICIPATING IN THE IIC EFFORT 19

Note: The main text of this whitepaper is contained in its first ten pages. The Appendices that follow contain information that is also available at the IIC web site, which is located at cfa-www.harvard.edu/~agoodman/IIC. A version of this document containing live web links is available at that site as well.

---

[1] A full list of IIC participants, and their affiliations, is given in Appendix E.

THE HARVARD INSTITUTE FOR INNOVATIVE COMPUTING

NOVEMBER 2004

Vision

The Institute for Innovative Computing (IIC) will make Harvard a world leader in the innovative and creative use of computational resources to address forefront scientific problems. We will focus on developing capabilities that are applicable to multiple disciplines, by undertaking specific, well-defined projects, thereby developing tools and approaches that can be generalized and shared. We will foster the flow of ideas and inventions along the continuum from basic science to scientific computation to computational science to computer science. We will train a next generation of creative and computationally capable scientists, build linkages to industry, and communicate with the public at large.

Implementation

The IIC will accept project proposals from groups of researchers at Harvard, potentially including IIC researchers themselves, whose scientific endeavor requires extraordinary computing-related capabilities. The proposals will vary in size, requiring anywhere from a few person-months to tens of person-years of collaborative effort. The IIC's projects will result in both new research in a traditional scientific discipline and new software, techniques, hardware, or systems that will be useful to future similar projects—at Harvard and beyond. The IIC will begin with a core staffing level of 30-40 at its inception as a virtual institute (~2005) and expand to ~80-100 at the time of construction of a facility, presumably in Allston, in ~2010. These IIC staff positions will be directly funded by Harvard University, although the IIC's mission and project palette will also expand through partnerships with government agencies and industry. The IIC staff will be comprised of both teaching faculty, who will teach and train both undergraduates and graduate students, and professional engineers and designers whose inventions and technical accomplishments will be valued equally with the scholarship of the faculty. The IIC will be organized into five branches: (1) Databases & Algorithms; (2) Hardware & Systems Integration; (3) Visualization; (4) Internet, Web and Grid Computing; and (5) a new Museum, on the grounds of the IIC, that would educate the public about the wide variety of science facilitated by high-end computing.

MOTIVATION

Computers have driven change in science over the past fifty years, and over the next fifty, science will drive changes in computing. The Institute for Innovative Computing (IIC) will facilitate scientific discovery by investing in scientific research that can, and will, only be done with the help of computers. The IIC will focus in particular on areas that cut across disciplinary boundaries.

Modern science is currently blessed by relatively newfound abilities to measure nearly "everything" researchers always wanted to know about key questions (e.g. sequencing the full human genome) on reasonable time scales, and to simulate basic processes not understandable through traditional pencil-and-paper calculation (e.g. the gravitational interactions of chunks of rock forming a Solar System). As a result, we are left with an embarrassing mismatch between how much data we can generate, and how much we have so far understood. Future discoveries will require close collaboration between the world of computer science and the scientific disciplines. Each, in their own right, has made remarkable progress, but the IIC is about making progress at the intersection of computer and other sciences. This realm of scholarship has fallen between the cracks in our existing organizational structure, as it lies between the traditional definitions of computer science and the life and physical sciences. The IIC is intended to bridge this gap, emphasizing the interdisciplinary applicability of innovative computing across the scientific frontier. Education is a central ingredient in our vision, in both the classroom and the community.

While many schools and institutions have recognized discipline scientists' need for high-performance computing, only a few groups[2] have realized—as we have—that people, and interdisciplinary collaboration, are more important than machines. The IIC strategy is to concentrate on adding a cadre of experts on the needs and techniques of scientific computing (not computer science) to the core Harvard staff, rather than on buying CPU cycles. The Director of the new Division of "Shared Cyberinfrastructure" at NSF, when informed of our group's IIC plans, said: "This is exactly what needs to be done, and Harvard will be seen as a great and visionary innovator if they do it first."

We outline a plan below to start the IIC now, with growth and planned evolution to an Allston-centered Institute in dedicated space by 2011. We believe that starting now will not only put us out ahead, but will also enable our ideas to be validated in practice, and to be improved by in-process learning and course correction.

EXPECTED BENEFITS OF THE IIC

Advancing Science with New Tools and Approaches

Collaborative research projects will be carried out at several of Harvard's schools, with faculty in (at least) Astronomy, Biology, Chemistry, Computer Science, Government, Earth and Planetary Sciences, Education, Environmental Science, Genetics, Neurology, Physics, Radiology and Statistics. We have identified a wide range of potential IIC projects (Appendix A), and four sample descriptions of these are provided in Appendix B.

Education

IIC members will offer new courses and research experiences to both undergraduate and graduate students, providing an opportunity for apprenticeship in data-intensive science that does not exist at Harvard, or almost anywhere, today.

The new IIC-hosted Museum, and its teacher-training programs, will engage the community at large and the K-12 sector with Harvard students, faculty and staff.

Collaborations and Linkages Across Harvard's Schools

The courses, seminars, and research opportunities offered by the IIC will provide a natural mechanism for the sharing of ideas, techniques and people between the diverse Harvard constituencies that will participate.

The IIC will complement and strengthen existing related activities, centers, and institutes at Harvard (see Appendix C).

Cooperative Endeavors and Knowledge Exchange with Industry

We expect to develop collaborations with industry very rapidly as we bring in exceptional scientists and technologists. These collaborations will take several forms, including: (a) Visiting Scientist program, targeting industrial partners interested in working with IIC faculty; (b) research collaborations to develop new methods, software, and instruments; (c) licensing of IIC technology to industry; (d) joint ventures.

---

[2] Among the very few but notable examples are: The Computation Institute at the University of Chicago (http://www-fp.mcs.anl.gov/ci/), the Caltech Center for Advanced Scientific Computing Research (http://www.cacr.caltech.edu/), and the UK National eScience Center of the Universities of Glasgow and Edinburgh (http://www.nesc.ac.uk/index.html).

Some companies we might initially expect to target as collaborators would be information technology, instrument, pharmaceutical and biotechnology companies, such as IBM, Intel, Apple, Microsoft, Hewlett-Packard, Mitsubishi, Oracle, Siemens, Biogen IDEC, and Novartis.

We also anticipate the possibility of developing joint ventures with leading media companies involved in animation, computer graphics and special effects.

Additional Benefits to the University

The IIC will provide a forum for innovation and communication. Our multidisciplinary seminars and courses will lay a foundation for bridging the boundaries between existing departments and schools. We fully expect that once the different IIC stakeholder communities become informed about the challenges and opportunities in other fields, innovative interdisciplinary work will follow. There is a real potential for the IIC to foster cross-cutting work at the multifaceted interfaces between traditional scientific fields and computing technologies.

We expect that the IIC will attract graduate students, postdocs, and senior visitors from both academia and industry who will enrich and contribute to our campus. The public outreach effort at the IIC will provide a mechanism for the University's Allston Initiative to directly benefit the communities and schools who will be our new neighbors.

What the IIC will *Not* Do

It is important that we be clear about what the IIC will not do. We do not intend to subsume or weaken existing Harvard computing efforts within groups, departments or schools. The IIC will not address the existing, and very serious, shortfall in support for research computing infrastructure (including system administration) on campus. We recognize and call attention to the fact that the University must increase its overall support for research computing in general, and we see the IIC as addressing *one important aspect* of this broader issue. It is also important to be clear that the IIC is not intended as the host for a "soft-money" research institute. While we do expect very substantial government and industrial investment in the IIC's projects, which will allow those projects' horizons to expand, it is our intention for the IIC's core educational and research mission to be carried out by Harvard-funded staff.

AREAS OF IIC ENDEAVOR

The IIC will be arranged organizationally into the five "branches" described on p. 8: (1) Databases & Algorithms; (2) Hardware & Systems Integration; (3) Visualization; (4) Internet, Web and Grid Computing; and (5) Museum & Outreach.

Here, we describe several central areas of work; the discussion of the IIC's structural organization as an institution begins on p. 7.

Deploying, Exploiting and Integrating Large Complex Databases

After extensive discussions, involving more than 50 people within the Harvard and Boston-area research community (see cfa-www.harvard.edu/~agoodman/IIC/meetings.html), it became very clear to all of us that the disciplines of modern science share a common set of database challenges that can only be addressed by advances in computation

Although large and complex data sets are growing rapidly, knowledge extraction tools remain relatively primitive. Many IIC projects will involve building tools to facilitate the access, analysis, and distribution of large data sets. Most of these tools will be useful in a cross-disciplinary way, since many problems are shared amongst disciplines. Sample shared challenges include dealing with multi-scale data, or data from a variety of observational, experimental, or clinical techniques, all of which are relevant in one investigation. Finding "relationships" or patterns in vast datasets—either purely empirically, or using models based on theoretical principles—is also a universal challenge.

The grand computational challenge represented by integrating data and knowledge in large science domains was described, as it relates to brain science, in a recent paper by the heads of five of the NIH Institutes:

> *"Biology has become an information science … The challenge now is to integrate this information into a coherent, accessible form that permits hierarchical analysis from RNA to protein to morphology to connectivity to function in a universal language while preserving fidelity."* Insel, Volkow, Li, Battey, & Landis (2003)

We intend to attack precisely this category of problem, which cuts across many scientific areas.

An important task of the IIC will be to develop effective means of sharing the database-handling innovations created there, and elsewhere, with the larger research community at Harvard. We expect that this knowledge will be disseminated in part through shorter-term collaborative projects where systems developed for one application are implemented in another field of research by IIC members.

Data Provenance

Extracting reduced data from large and evolving data sets, using software that is also evolving with time, is a complex problem. Tracking which data were used in conjunction with what code version to produce a given representation of reality is an unsolved problem. Scientists are often unable to simply reproduce reduced data sets. The IIC will address this issue by developing self-describing data structures to track and log the provenance of reduced data sets.

Data stored and mined from databases must have descriptors of experimental methods, calibration, and inherent uncertainties attached. This information is part of the "metadata" needed to use the data (e.g. knowing the total number of patients involved on clinical trials on a drug, and knowing where those trials took place).

In some cases (e.g. astronomy) keeping careful track of data's source is key, whereas in others (e.g. human subject research), obscuring ("anonymizing") the data's source (while maintaining a secured provenance record at the core of the system) is the goal.

Visualization, Communication and Design

Well-designed images and graphs provide unparalleled insight into physical, chemical, and biological interactions and processes. The IIC will investigate and implement novel solutions to visualization and communications challenges, including completely new graphical approaches to display of complex information, made possible through advanced computation. It will approach visualization and communication challenges, in collaboration with discipline scientists, with the objective of creating novel, accessible and informative graphics for communicating scientific results both within the scientific community, and to the general public. In many cases, the most successful visualizations of data requiring high-performance computing to create will be accomplished through collaboration with psychologists or visual artists.

Specific visualization and communication challenges common to most disciplines include:

Escaping Flatland: Displaying multi-dimensional, multi-scale images on a page or screen (e.g. how to display a three-dimensional brain tumor's growth, on a computer screen, when it has been imaged over three orders of magnitude in scale).

Designing and Pushing the Limits of Image Processing: How to speed software and hardware enough to handle "all" the data to be had. The high-end computing hardware discussed in the numerical simulation section below will be important, in addition to new expertise, in these efforts.

Truth in Imaging: Addressing the important questions of *truth* in data representation—what does false color really mean…how transparent should we "paint" an object? How can we create a system of tracking and archiving those manipulations?

Representation of Uncertainties: Uncertainties are represented graphically in different ways in different fields, and seldom portrayed at all for the public. As data representations become more "interpretive," thanks to computer processing, the proper tracking of both the data manipulation, and its intrinsic uncertainty, becomes critical to scientific interpretation.

Redefining "Communication": Going beyond what is now accepted as graphical representation and delving into the challenge of how to *really* make data communicative and accessible, especially at discipline boundaries, and to the public.

Large-Scale Distributed Scientific Collaborations via the Internet

Large, geographically dispersed, scientific collaborations, with distributed computation and data, are increasingly conducted over the Internet. The Human Genome Project may have been the prototype for this kind of joint work – it simply could not have been accomplished except as a distributed project with the cooperation of widely dispersed centers of expertise and technology. Such large-scale collaborations over the Internet, often known as *"e-Science,"* raise a host of issues, including distributed access to data, efficiently distributed computation, pooling of remote computer resources, and integration of knowledge (metadata) systems based on primary data.

IIC stakeholders involved in these large collaborations are presently frustrated by the lack of relevant e-Science tools. There are many areas where the IIC will likely strive to improve the flow of data and ideas in research consortia, but one sure focus is on "grid" computing.

Grid computing allows large scientific collaborations to pool their computing resources and data to form an effective computing system across a wide area network. The "SETI@Home" screensaver project is a rudimentary (but brilliant) example of a grid. Many applications in physics, astronomy and biology are now employing homegrown grid computing, but working, user-friendly, grids are still in the development stage. While Harvard is at the frontier of many of the scientific disciplines that will require grid computing, it is presently falling behind in grid technology. The IIC will allow us to develop the expertise and support in grid computing that the Harvard research efforts requiring it need to remain world leaders.

Theoretical Modeling and Simulations

Numerical simulations are an increasingly important component of modern science. Protein folding, the evolution of the Earth's magnetic field, or the formation of structure in chaotic gravitational systems cannot be fully studied and understood with pencil-and-paper analytic calculation. Most simulations are *not* well suited to running in the same kind of distributed (e.g. grid) computing environment as are data handling or reduction tasks. Instead, they require high-speed machines currently comprised of many processors connected together as seamlessly as possible. The interaction of simulation code and these machines' performance ultimately determines how much science can come from any particular calculation. Multiple branches of the IIC (see p. 8) will work together on projects aimed at optimizing the process of numerical simulation. And, the IIC will host some of the only computing hardware at Harvard suitable for these numerical experiments. Perhaps most importantly, the IIC will provide a forum for the scientifically-diverse, but practically-linked, simulation communities to share and refine their tools, techniques and insights.

Integration of Software, Hardware and Networks

The IIC is not envisioned to be large enough to contain a full-scale hardware development lab, where components are fabricated from raw silicon wafers or even individual chips. That said, building innovative experimental hardware systems out of off-the-shelf components is often the most efficient way to solve scientific problems. What is largely lacking at Harvard today is the expertise to know which kinds of hardware, using what kind of new technology, being developed by what company, might help with a particular computational challenge. Sample technologies to be explored at the outset will include: inexpensive data-acquisition devices ("sensors"); new interaction devices for visual exploration; and the use of graphics processing units as co-processors for scientific computing.

In typical applications, hardware is not a challenge on its own. The *interaction* of code, hardware, and the network on which both are deployed needs to be optimized. Work that stretches the limits of components and seeks to dramatically improve these interactions is an emerging academic discipline that will be highlighted at the IIC.

The group in DEAS headed by J. Sircar, currently focused on just this kind of software-hardware-network integration, will serve as a repository of knowledge and expertise on: networking, interoperability; grid computing; large data storage; code optimization; and other topics in high-performance computing systems. The DEAS will also host the IIC's hardware, which will include a high-performance computing cluster, before the IIC complex is constructed.

Education within Harvard

To succeed in its mission, the IIC needs to train a new generation of creative and computationally-capable scientists and scientifically-educated computer scientists. This requires a fundamentally new set of skills that neither of the disciplines (sciences or engineering) currently teach.

Some of the initial hires of the IIC will be of teaching faculty (see below). These faculty will teach *new undergraduate and graduate courses* focused on "data-intensive science." In addition, all IIC members will be happy to involve students in the IIC's research efforts, through the same kinds of research assistantships currently in place at most of Harvard's schools. We will allocate a portion of the IIC's high performance computing, as a "sandbox," to support these educational goals.

The scientific visualization tools developed at the IIC will also facilitate creation of new modes of teaching Harvard undergraduates in both science *and* non-science fields. The recent report on the Harvard College Curricular Review highlighted a growing need to better educate both future scientists and non-scientists. As part of an effort to address this challenge, faculty from the FAS Division of the Life Sciences have launched projects that seek to teach science in a newly synthetic fashion beginning in the freshman year. Besides new courses in a revised undergraduate curriculum, these efforts also involve the development of clustered simulations and other multimedia where students explore the connections between multiple fields in the context of a compelling issue or question. The resulting synthesis of the life sciences with the physical sciences and math confers a perspective on science and technology that should be an essential ingredient of any undergraduate education in the 21st Century. In addition to the visualization tools, the IIC would serve as an arena for the development and testing of simulations and the software architecture to support these curricular efforts. Therefore, we expect the rich cross-disciplinary research supported by the IIC would be linked through the institute with the ongoing evolution of undergraduate teaching at Harvard.

In addition to educating students, IIC members plan to educate *each other* and the Harvard research community at large. The IIC, even before its construction as a physical institute, will host weekly seminars on somewhat specialized topics, and monthly colloquia on more broadly-applicable work. The IIC is also likely to facilitate the organization of both internal (to Harvard) and external workshops on topics in innovative scientific computing.

Outreach: Education beyond Harvard

As explained below, in its brick-and-mortar form, the IIC will host a new Harvard Museum, showcasing the kinds of science done at the IIC.

Before the Museum, which will serve as a natural focal point for the IIC's outreach programs, is constructed, we plan to begin a new kind of educational effort centered on K-12 teacher education, and involving high-school students in IIC projects. Much of the science taught in schools today is a bit out of step with the science being done in cutting-edge research groups. The IIC will offer an opportunity to bring teachers to Harvard to literally show them the new science, and to teach them about some of the techniques used to accomplish this science. Scientifically and mathematically gifted high-school students (a very underutilized resource) will be directly involved in IIC projects, both remotely and at Harvard, through open-source code development and grid-based computing projects. We have been working closely with the 60-person Science Education Group centered at the Harvard-Smithsonian Center for Astrophysics on outreach planning, and we are sure that the IIC's educational programs will fulfill a completely unmet need.

IIC Staffing and Organizational culture

The IIC is intended to produce a large body of discipline-transformative research and inventions: applied, interdisciplinary research in which computational science is developed to enable other research domains, which in turn drive applications and developments in computational science. What we are attempting to create is a new kind of information ecology at Harvard, based upon communities of practice that cut across traditional boundaries and intersect within a space of shared problems in computation. This effort requires strong support in IIC's organizational culture, which will emphasize collaborative practice and invention.

Staff will be recruited both from the ranks of highly accomplished academics, and from senior experts whose careers have been primarily in industry (e.g., Microsoft, IBM, Intel, Apple, Pixar, Siemens, Sony, etc.). Criteria for promotion will give equal weight to scholarly activities, and to technological invention. There will be no distinction made in status between teaching and non-teaching faculty; or between scientists and engineers (or for that matter artists and designers working in the visualization program).

Hiring levels and positions will be modeled after the system at Harvard Medical School, where, again, there is no distinction made between clinical and non-clinical faculty. Appointments will be for a term of years dependent upon level, and renewable for the life of the program (except visiting appointments). In some cases, collaborating with a particular outside expert who may wish only to visit, but not move to, Harvard will be the best course, and we plan on fostering national, international, and industrial collaborations with an active visitor program.

A list of planned position categories is as follows:

Term	Teaching Positions	Research Scientists	Designers or Specialists
Weeks to months	Visiting Scientist	Visiting Scientist	Visiting Designer
2 to 3 years	Postdoctoral Fellow	Postdoctoral Fellow	Postdoctoral Fellow
1-year, renewable	Instructor	Scientist	Designer
3-year, renewable	Assistant Professor	Senior Scientist	Senior Designer
5-year, renewable	Associate Professor	Principal Scientist	Principal Designer
Appointment does not need renewal	Professor	Distinguished Scientist	Distinguished Designer

The organizational culture will value excellence in one's discipline, as reflected in high-quality interdisciplinary collaboration with deep impact upon the field of endeavor.

The proposal process will drive the detailed breakdown/allocation of staff between branches and seniority levels. However we envision the initial staffing distributed across these layers from the "Professor" level to "Instructor" level in approximately the ratio 1: 2: 3: 3, as a rough guide, with ~30-40 staff hired in the first Phase. A guide to planned growth is given below.

The IIC Director will be appointed by and will report to the Harvard Provost. The Director will bear overall responsibility for the activities of the Institute, will lead development of its scientific vision, goals and strategy, and will assure its program is strategically aligned with collaborating disciplines and activities across Harvard. S/he will be supported by an Executive Director, who will have principal responsibility for implementation of the vision and goals as a coherent, well-managed set of projects and technology collaborations, and for overall operations of the Institute.

An External Advisory Committee (EAC) will advise the Director, and provide the Provost with an annual assessment of the IIC's performance and program quality. A Program Advisory Committee (PAC) will evaluate and rank projects and initiatives proposed to the IIC. The PAC, to be appointed by the Director, will be comprised primarily of Harvard researchers plus a few external members, whose expertise reflects the broad interests of the IIC stakeholders within the Harvard community. The Director will allocate 75% of the IIC resources (primarily the time of collaborative personnel) in accord with the recommendations of the PAC. The remaining 25% will be placed at the discretion of the IIC Director to seed new projects, to undertake high risk/high payoff endeavors, or to support other work that the Director deems meritorious.

The PAC will assess proposed IIC projects based upon their scientific strength, the likely broader impacts of the project, and the match to the IIC skills, experience and growth goals. At the time a project is begun, a "Project Manager" will be identified from amongst the project's collaborators (comprised of both discipline scientists and IIC members). This Manager, who may be hired specifically for a particular large project, will be responsible for the project's execution and for clearly identifying metrics for tracking progress and performance. We expect that the management structure and the complexion of the IIC program will evolve over time, so we intend to implement an organizational structure that can readily adapt to new needs and opportunities.

PLANNED STRUCTURE & EVOLUTION

The IIC will implement a bold, interdisciplinary vision in scientific computation at Harvard. We envision an Institute divided into five main areas of expertise: (1) Databases and Algorithms; (2) Hardware and System Integration; (3) Visualization; (4) Internet, Web and Grid Computing; and (5) Museum and Public Outreach. Many projects will be implemented across these areas of expertise: for example, a project on semantic grid computing for education might draw on all five areas.

The IIC will begin as a Virtual Institute linked together via weekly seminars and colloquia, by significant shared computational infrastructure, by web technology and other approaches specifically designed to enable social knowledge networks. The visitors program will begin at the outset to establish strong exchange and links with other institutions and with industry.

We envision three major phases of growth, with the final phase hosting a transformation in which the Institute becomes centered at a dedicated custom-built facility in Allston. This facility will not only become a vibrant and essential part of the Allston campus, but will serve as a nexus of collaboration tying together many disciplines across the entire Harvard campus. (In social network analysis, the IIC might be termed a "boundary-spanning node".)

The "rolling start" as a Virtual Institute will enable Harvard to establish a strong leadership position among its peers, some of whom have also begun to realize the importance of the ideas discussed in this whitepaper. The phased-in approach will enable our concepts, which are new and transformative, to undergo stepwise refinement in practice as we grow toward the Allston environment.

Planned evolution is as follows:

Phase 1 (2005-2007)

- External Advisory Committee appointed
- IIC Director and Executive Director hired
- Pilot projects selected by a Program Advisory Committee
- "Startup space" secured for leadership and core staff
- Program leadership and core staff hired, begin initial work on selected pilot projects in collaboration with other Harvard researchers participating in the IIC's programs
- ~30-40 researchers and staff hired over two years, to accomplish Pilot projects
- Hires will include at least three new faculty, whose work focuses on the IIC mission
- New undergraduate courses, and new graduate training program in "data-intensive science," established, with input from new faculty
- Establishment of a Visitors Program for short-term collaborations with outside experts, to be administered by the IIC Director
- Applications for government funding to expand IIC efforts

Phase 2 (2008-2010)

- Begins with critical evaluation of Phase 1 by External Advisory Committee
- Begin planning for new Museum, to be constructed in Phase 3
- In conjunction with Museum planning, begin to develop and implement teacher-training materials & new educational programs
- Expanding in "successful" areas, and contracting in others, staffing level expands to ~60-75. Some of the expansion includes new kinds of researchers and staff to work on educational programs.
- Planning for construction to take place ~2010.

Phase 3(2011 and beyond)

- Completion of a physical institute, organized into five branches, if the same five still make sense
- Staff expansion, to ~80-100, expanding in successful areas, and adding more museum/education staff (some positions expected to be funded by outside sources)
- Opening of new Museum
- Founding of any new "Departments" within the schools at Harvard, if appropriate

We feel that the IIC vision we have put forward here can and will lead to a deep transformation of Harvard's practice in research computing. It will radically enable a host of computationally-enabled disciplines in a synergistic way. It will serve as a vibrant nexus of collaboration tying disciplines together. And it will be a powerful attractor, binding many Cambridge-, Longwood- and Charlestown-based departments of the University to colleagues in Allston and to each other. It will enable Harvard to establish national and international leadership in emerging disciplines with profound impact upon society.

IIC Leadership: Start-up Tasks and Approach

We propose a rapid startup of the first phase of the IIC during Q1-Q4 2005. During this phase, the IIC will be a semi-virtual institute, with many of its members spending most of their time working with existing groups around the campus, in their existing space. However, we envision the need, within the IIC's first year, for a "home base" suite of about a dozen contiguous offices in Cambridge to house its leadership, and core and administrative staff. Roughly speaking, Q1 of 2005 would be a transition from our current Planning stage to a Launch stage in Q2, leading to an Early Ramp-up stage in Q3-Q4. By the end of 2005 we would expect all critical leadership roles to be filled, and a first round of (pilot) projects to be running at some reasonable initial level of effort.

Activities Ramp-up would continue through 2007, the end of the Phase 1. This year would also be a year of consolidation and re-direction, with many substantial project completions, a comprehensive program review, and planning for the next three years of IIC's existence and growth.

Beginning in Q4 2004 and ending in early Q1 2005 we propose to prepare a detailed business plan, budget, and launch schedule for the IIC; begin discussions on the initial set of pilot projects, firm up our External and Intramural Advisory Boards, and secure first-year funding. In Q1 2005 we would hope to finalize the selection of pilot projects, create the hiring plan, work to secure "start-up" space for core staff, and begin recruiting.

We hope to be able to bring in a start up leadership team consisting of Director and Executive Director, supported by some senior administrative positions (e.g. Finance Officer), beginning in Q2 2005. This startup team would be responsible for the rapid and effective launch of IIC as a semi-virtual institute and its transition to permanent status. Some leadership positions in this team - for example, the Director - might be specifically designed as temporary "on-leave" assignments for the sole purpose of launching IIC; while others might continue through the post start-up period. Some positions could also very well be part-time, or consulting roles

The Director and Executive Director would lead an aggressive recruiting process to hire heads of most of the five internal groups beginning in late Q2 - Q3 2005. These Group Leaders would then recruit staff in their areas, with extensive mutual consultation. Recruitment will be guided by a combination of three factors: (a) long-term expertise and capabilities required by the Institute, (b) near-term project and start-up needs, and (c) availability of outstanding individuals. Clearly, the ability to bring in effective start-up leadership at an early stage would be a key to early launch. We have identified viable options for accomplishing this task, which, however, do require some further discussion and consideration.

All of the initial IIC hiring will be carried out in close collaboration with a select group drawn from the full list of IIC stakeholders in Appendix E, and the process will include several of the authors and editors of this document.

This plan is aggressive, but we believe it is highly achievable, based on the level of commitment, mutual agreement, and preparatory work already in place; and the great enthusiasm for this project across the Harvard Community.

APPENDIX A:
A REPRESENTATIVE SAMPLING OF IIC PROJECTS

Note:

A longer, continually evolving and more complete list of suggested IIC projects is online at:
http://cfa-www.harvard.edu/~agoodman/IIC/documents/whitepaper/one_pagers.html

Short descriptions of the four projects marked by (*) are included in Appendix B.

IIC Project Titles

The Semantic Web for Neuromedicine: An Industry/Academic/Non-Profit Collaboration Using Rapid Evolutionary Prototyping – *Tim Clark* (HMS/MGH)

EScience in Translational Biomedicine: a New Model for Interdisciplinary Knowledge Transfer and Organizational Memory in Biopharmaceutical R&D – *Tim Clark* (HMS/MGH)

Modeling, Manipulating, and Visualizing N-Dimensional Data – *Martin Elvis* (CfA)

Inferring the Real Universe: Overview – *Pepi Fabiano* et al. (CfA)

The Harvard-CfA Astronomy Portal – *Pepi Fabiano* et al. (CfA)

Parsing, Displaying, and Serving the COMPLETE Database on the Web: A Stepping Stone to the Virtual Observatory – *Alyssa Goodman* (CfA/FAS)*

Next Generation Imaging in Radio Astronomy: Visualization of the Earliest Structures to Form After the Big Bang – *Lincoln Greenhill* (CfA)

Next Generation Imaging in Radio Astronomy: Construction, Mining, and Serving of Instantaneous All-Sky Images – *Lincoln Greenhill* (CfA)

Analysis of Electroweak Symmetry Breaking at Harvard and Resource Allocation on Grids – *John Huth* (FAS)*

The Harvard Virtual Brain – *David Kennedy* (HMS)*

The Syntuit Project Using Multimedia to Achieve Synthetic Intuition in the Undergraduate Life Sciences – *Robert Lue* (FAS)*

Inference in a Cellular Web – *Fritz Roth* (HMS)

A Tool Box for Data Manipulation, Management and Visualization in Genomic Research – *Eitan Rubin* (FAS)

The Private Universe of Visual Imagery – *Matt Schneps* (CfA)

Annotation of DNA Sources – *Brian Seed* (HMS/MGH)

Visualizing Information Sources – *Brian Seed* (HMS/MGH)

Multiscale Biomedical Imaging Tools and Databases – *Stephen TC Wong* (HMS/BWH)

APPENDIX B: ONE-PAGE DESCRIPTIONS OF SAMPLE IIC PROJECTS

In preparation for the creation of this whitepaper, we held several meetings, bringing together researchers from across the Harvard community, along with a few local outside experts, to discuss the IIC. Participants in these meetings were asked to submit one-page descriptions of realistic sample projects they might propose to the IIC. We received 32 short project descriptions, a partial list of which is given in Appendix A, along with a link to the online descriptions of all the projects. Here, we highlight just four projects selected to show the range and level of work in which the IIC members will engage.

THE HARVARD VIRTUAL BRAIN

David N. Kennedy
Departments of Neurology and Radiology, Harvard Medical School

Harvard University possesses a vast neuroscience resource. The diversity of this information includes: molecular data (from genes to cell specific expression), neuroanatomy (from the ultrastructure of synapses to systems level anatomy), electrophysiology (from single channels to scalp surface EEG), brain imaging (from structure to function using a wide variety of imaging techniques), behavior (from basic sensory operations to complex cognitive tasks), species (from mouse, through primate to human), and neurochemistry (from neurotransmitters to receptor binding sites). As research in neuroscience continues to advance, the capacity of the individual investigator to retain all relevant information is severely taxed.

The distribution of neuroinformation within the Harvard Community is highly dispersed. At an overly simplistic level, HMS and FAS host vast quantities of basic neuroscience knowledge; FAS hosts vast quantities of behavioral and cognitive information, HMS and the Affiliated Hospitals host vast information in the applied clinical domains of neuroscience, DEAS supports a substantial array of tools, analysis methods, databases and computational infrastructure. In reality, these domains of information and their 'locations' in the Community are inextricably intertwined; but at the same time virtually unconnected. The development of a Harvard-wide neuroinformatics effort that links each of these scales and classes of information will be critical to advancing our understanding of the brain across these scales and domains of knowledge.

The Biomedical Informatics Research Network (BIRN: www.nbirn.org) has developed many tools and specifications for the development of co-laboratories. The Harvard Virtual Brain will leverage this experience in order to encourage Harvard-wide collaborations among members of diverse research areas and scientific domains that traditionally conducted independent investigations and that fosters distributed collaborations in biomedical science by utilizing information technology innovations. Accomplishing this will include development of the necessary cyberinfrastructure that enables sharing and collaborative use of distributed biomedical databases and data collections, analysis and modeling software, and visualization tools, and creation of a scientist-friendly point of entry or "portal" to access and provide training on these tools. These tools will be used to implement a secure, federated data sharing environment that presents biological data held at geographically separate sites as a single, unified database. The overarching goal is to establish tools practices to enable the aggregation of data from virtually any laboratory's research program and assemble knowledge derived from disparate domains into a commonly accessible platform.

Key constituent technologies include: Assessment of Anatomic Structure - participants are examining neuroanatomical correlates of behavior, function, development and pathological insult, through large-scale analyses of patient population data acquired and pooled across sites; Assessment of Function - bringing together researchers in different aspects of functional neuroimaging to apply recently developed multi-modal and interdisciplinary techniques to investigate the neural substrates of behavior and function; and Assessment Across Scales - studying animal models of disease at different anatomical scales to test hypotheses associated with human neurological disorders. to share and analyze multi-scale structural and functional data and ultimately to integrate them with genomic and gene expression data. Each of these domain areas require consideration of methods for calibration and standardized reporting of observation, refinement of visualization and analysis tools, and implementation of federated, distributed database and knowledge management methods.

We envision two potential members of the IIC who would work together in making the Harvard Virtual Brain dataset accessible to all on the Web. One IIC collaborator would be expert in the design and searching of large multi-dimensional databases, and the other would be more focused on designing tools to use existing and new visual interfaces to access and analyze the database, through implementation of the Harvard Virtual Brain Portal. We expect that a total of six person-years (three years per participant, in parallel) would be needed to enable this interoperative environment of data sharing and knowledge management to be implemented. The result of this IIC work would be a new data retrieval/display system to accumulate information on the structure, activity, and function of the brain, in health and disease to actually solve the question of "How the brain works?"

PARSING, DISPLAYING, AND SERVING THE COMPLETE DATABASE ON THE WEB: A STEPPING STONE TO THE VIRTUAL OBSERVATORY

Alyssa A. Goodman
Department of Astronomy, Faculty of Arts & Sciences, Harvard University

The COMPLETE Survey (cfa-www.harvard.edu/COMPLETE) currently underway is the largest, most diverse, systematic, multi-wavelength study of star-forming regions ever undertaken. Star formation, in spite of its central role in the evolution of the Universe, is currently very poorly understood, and the COMPLETE database will be used to answer a tremendous variety of physical questions, including ones like "How did the Sun form?". Without systematic surveys like COMPLETE, the data available are just too statistically sparse and poorly archived to make headway on these questions.

The COMPLETE data set includes maps made at wavelengths from optical through radio, some of which have an added "velocity" dimension beyond the three dimensions (position-position-intensity) astronomical images usually contain. Each of the mapping techniques employed produces a unique "kind" of data that is ordinarily dealt with almost exclusively by experts on that "type" of data, and very few researchers are expert in all the techniques. The tools for displaying and interpreting these kinds of data are, at this point in time, technique-specific. Trying to retrieve, use, and display COMPLETE and other similarly diverse astronomical data now is akin to trying to assemble a car from all of the one-thousand needed parts, but using directions partially in Chinese, partially in English, and partially in French, with some illustrations that can only be viewed on Windows PCs, others only on Macs, and some others only with the right brand of 3D glasses. Making all of COMPLETE's data readily searchable and accessible to astronomers worldwide, regardless of their expertise or computing platform, is an unprecedented—but definitely achievable—challenge in both database design and visualization.

We envision two potential members of the IIC who would work together in making the COMPLETE dataset accessible to all on the Web. One IIC collaborator would be expert in the design and searching of large multi-dimensional databases, and the other would be more focused on designing tools to use existing and new visual interfaces to access and analyze the database. We expect that a total of four person-years (two years per participant, in parallel) would be needed to create a streamlined working system.

The result of this IIC work would be a new data retrieval/display system that would be the first to handle what is known as "velocity resolved" or "spectral line data cube" data sets that intrinsically have four dimensions. This system would be incorporated into all of the Virtual Observatory interfaces now being developed (see http://www.us-vo.org/) to provide this kind of access to multi-wavelength, multi-dimensional data sets across astronomy.

THE SYNTUIT PROJECT: USING MULTIMEDIA TO ACHIEVE SYNTHETIC INTUITION IN THE UNDERGRADUATE LIFE SCIENCES

Robert A. Lue
Department of Molecular & Cellular Biology, Faculty of Arts & Sciences, Harvard University

A great challenge facing Life Sciences education today is the development of more effective ways of teaching students to integrate concepts from multiple fields so as to gain a truly synthetic perspective on biological issues. As the allied fields move away from a primarily reductionist understanding of biological phenomena, the need for this kind of synthetic intuition has grown both in the classroom as well as in research. As the newly formed Life Science Curriculum Committee in FAS embarks on a major revision of the Harvard undergraduate curriculum in the life sciences, one of the major goals is to offer students a more integrated introduction to science beginning in the freshman year. The Syntuit Project seeks to support this kind of synthetic learning by clustering different types of educational multimedia together with diagrams, still photography and text, all of which address different but inter-connected aspects of a particular life sciences topic. These Syntuit Modules (SMs) will be built into an interactive graphical user interface that synchronizes and hyperlinks information across all media types. Thus, the user will be able to seamlessly connect the research data behind the particular model or simulation while also pursuing connections with other allied disciplines. SMs will be generated for topics spanning all the departments in the FAS Division for the Life Sciences. In addition, the project seeks to develop a personnel infrastructure for the sustained development and revision of such media modules as an integral part of Harvard's teaching mission. Although it is beyond the scope of this white paper, a significant portion of the Syntuit Project involves the creation of authoring and development communities composed of faculty, post-doctoral fellows, graduate and undergraduate students together with multimedia professionals and programmers.

Syntuit Modules and the Syntuit User Interface

The kinds of media that are incorporated into SMs include three-dimensional animations that place molecular structures and events in a cellular or physiological context; two-dimensional animations that illustrate sequential processes and the relevant underlying chemistry; interactive macromolecular structure tutorials based on atomic coordinates; digital video clips that show important research techniques and interviews with scientists discussing the history, current use and future directions of a given method. Pilot examples of each type of multimedia are already under development as part of the BioVisions Project in the Department of Molecular and Cellular Biology. We envision collaborations with members of the IIC from the Visualization Institute and the Database & Algorithms division. In the former case, the collaboration would seek to develop innovative ways of visualizing molecular processes that support conceptual integration into a broader biological context. In addition, the VI collaboration would explore new ways of synchronizing content across different types of media and within a unified interface. Significant participation from a range of Life Sciences faculty will also drive the establishment of a unified visual language for modeling common processes that will serve as a fundamental visual vocabulary for future simulations.

We envision the closely allied D&A collaboration as focusing on the development of algorithms and a software architecture that will directly link biological simulations to quantitative data and mathematical models. This combined with new methods for annotating and synchronizing animation/digital video to text and still images will form the computational backbone of the Syntuit User Interface (SUI). In addition, the development of simple authoring tools that allow instructors to cluster and synchronize disparate media will ensure the sustained adaptation of SMs for use in different undergraduate courses, research settings, and at the pre-college level. The development of such a software toolbox will open up the SUI to instructors who can then customize and create their own modules from any media to which they have access.

ANALYSIS OF ELECTROWEAK SYMMETRY BREAKING AND RESOURCE ALLOCATION ON GRIDS

John Huth (Physics), Margo Seltzer (EECS), Joy Sircar (DEAS)
Faculty of Arts & Sciences, Harvard University

Recent computing grid activities in the U.S. have demonstrated the viability of large scale computational grids via the "Grid3" persistent test bed. This has linked some 30 computing centers for over 10 applications that run concurrently, and ship 20+ Terabytes per day over the network. This grid work has served the needs of a number of scientific communities, including the experiments at the Large Hadron Collider Experiment, ATLAS.

The ATLAS experiment is designed to explore the origins of the symmetry breaking in the weak and electromagnetic forces. It will acquire Petabytes of data per year and the reconstruction and analysis of the data will be done by a distributed collaboration of over 2000 physicist. Within the Boston area, a consortium of five universities (Harvard, BU, Tufts, Brandeis and MIT) have concentrated on the construction of the muon subsystem, and are building capabilities for the analysis of data from the experiment. Harvard has played a leading role in the development of computing for the ATLAS experiment and the use of grid computing in sciences.

The goal of this proposal is two-fold: 1.) is to provide a premier facility for the analysis of data from the ATLAS experiment, and 2.) to create an important test bed that will attract users to investigate important problems in resource allocation in a grid environment. This proposal will build on two existing ITR proposals – one is for Prof. Huth that emphasizes using LHC applications to create "resource usage predictors", the second is a joint BU-Harvard proposal involving Prof. Huth and Prof. Seltzer to investigate the creation of an economic basis for resource allocation on computing grids.

Up to now, most of the resource allocation on grids has been done "by hand" (i.e. by a small number of users). This is intrinsically not scalable to large numbers of users, and cries out for another mechanism for resource allocation. An obvious approach is to impose a kind of "resource economy" on the grid users to allow resource allocation and scheduling.

We propose to leverage the existing ITR proposals from the NSF and complement this effort with resources from the IIC to provide a facility that is useful for both resource allocation techniques and also to provide the basis for data analysis from the LHC. The two pieces that are required are personnel in the form of a graduate student, and a higher-level computer-literate scientist to implement the resource allocation and prediction policies in the context of the ATLAS experiment as the driving application. Coupled with this, we propose the creation of a 100 Terabyte disk-farm. The disk farm is sufficiently large to be an "attractive" influence to get a number of users who will want to access the data collected by the ATLAS experiment. This work will be carried out coherently with the other Boston-area universities. In particular, this will be the basis of a coordinated proposal with Boston University to establish an analysis center for LHC data, and provides a substantial resource for both experimental and theoretical physicists at Harvard.

APPENDIX C:
SAMPLE RESEARCH INSTITUTES AND INITIATIVES AT HARVARD WHOSE MEMBERS WOULD COLLABORATE WITH THE IIC

The Bauer Center for Genomics Research (CGR)

The Center for Imaging and Mesoscale Structures (CIMS)

Harvard Center for Neurodegeneration and Repair (HCNR) (and associated "Cores")

The Harvard-MIT Data Center (HMDC)

Harvard-Partners Center for Genetics and Genomics (HPCGG)

The Harvard-Smithsonian Center for Astrophysics (CfA)

The Martinos Center for Biomedical Imaging (Martinos Center)

MassGeneral Institute for Neurodegenerative Disease (MIND)

Surgical Planning Laboratory (SPL)

APPENDIX D: OUTSIDE EXPERTS WHO HAVE AGREED, IN PRINCIPLE, TO SERVE ON THE IIC'S EXTERNAL ADVISORY COMMITTEE

ZINA BIN MILED, *Assistant Professor of Electrical and Computer Engineering*, Purdue School of Engineering and Technology, Purdue, Indiana

NEIL DEGRASSE TYSON, *Astrophysicist/The Frederick P. Rose Director, Hayden Planetarium, Department of Astrophysics*, American Museum of Natural History, New York, NY

DEBORAH ESTRIN, *Director, Center for Embedded Networked Sensing (CENS), Professor of Computer Science*, UCLA Computer Science Department, Los Angeles, CA

CAROL GOBLE, *Professor of Computer Science*, University of Manchester, UK

JIM GRAY, *Distinguished Engineer in Microsoft's Scaleable Servers Research Group and manager of Microsoft's Bay Area Research Center (BARC*), Microsoft Corporation, San Francisco, CA

JAMES HENDLER, *Professor of Computer Science*, University of Maryland, College Park, MD

SANG KIM, *Software and Tools for High-End Computing Division Director*, NSF Division of Shared CyberInfrastructure, Arlington, VA

RANDALL LANDSBERG, *University of Chicago, Cosmology Outreach Director*, Chicago, IL

MIRON LIVNY, *Professor, Computer Sciences Department* University of Wisconsin, Madison, WI

PATRICIA MONGER, *Technical Director, Research and High-Performance Computing Support Department*, Hamilton, Ontario, Canada

PETER QUINN, *Astronomy Professor*, European Southern Observatory, Germany

SYLVIA SPANGLER, *Program Director, NSF Science and Engineering Informatics Cluster, Division of Information and Intelligent Systems*, Arlington, VA

MARK SUBBARAO, *Astronomer, University of Chicago & Adler Planetarium & Astronomy Museum*, Chicago, IL

DINOJ SURENDRAN, *Data Visualization Specialist*, University of Chicago, Chicago IL

APPENDIX E: FULL LIST OF LOCAL RESEARCHERS PARTICIPATING IN THE IIC EFFORT

TENLEY ALBRIGHT, *Lecturer on Medical Education*, HMS

CHARLES ALCOCK, *Director, Center for Astrophysics*, CfA/FAS

JEREMY BLOXHAM, *Professor of Geophysics*, FAS (EPS)

COLLEEN CAVANAUGH, *Professor of Organismic and Evolutionary Biology*, FAS

GEORGE CHURCH, *Professor of Genetics, Director, Center for Computational Genomics*, HMS

TIM CLARK, *Director, Center for Interdisciplinary Informatics, MassGeneral Institute for Neurodegenerative Disease, Instructor in Neurology, Harvard Medical School,* MGH/HMS

THOMAS DEISBOECK, *Director, Complex Biosystems Modeling Laboratory*, CBML MIT

AARON ELLISON, *Senior Ecologist and Senior Research Fellow, Harvard Forest*

MARTIN ELVIS, *Astrophysicist, Chandra X-ray Center*, CfA

PEPI FABBIANO, *Head of the Chandra X-ray Center Data System Division*, CfA

BRUCE FISCHL, *Assistant Professor, Harvard Medical School*, HMS/MGH

FELICE FRANKEL, *Research Scientist*, MIT

ALYSSA GOODMAN, *Professor of Astronomy*, CfA/FAS

ROY GOULD, *Science Education Specialist*, CfA

LINCOLN GREENHILL, *Senior Scientist*, CfA

MICHAEL HALLE, *Instructor of Radiology, Director of Technology Development*, BWH/HMS

LARS HERNQUIST, *Professor & Chair of Astronomy*, CfA/FAS

JOHN HUTH, *Professor & Chair of Physics*, FAS

ADRIAN IVINSON, *Professor of Genetics, Director, Harvard Center for Neuro-degeneration & Repair*, HMS

DIAB JERIUS, *Astrophysicist, Chandra X-ray Center*, CfA

EFTHIMIOS (TIM) KAXIRAS, *Professor of Applied Physics & Professor of Physics*, FAS

GARY KING, *David Florence Professor of Government, Director of Harvard-MIT Data Center*, FAS

TOM KIRCHHAUSEN, *Professor of Cell Biology*, HMS

MARC KIRSCHNER, *Chair, Department of Systems Biology*, HMS

DAVID KENNEDY, *Associate Professor of Neurology, Harvard Medical School, Assistant in Neurology (Neuroimaging), MGH, Co-Director, Center for Morphometric Analysis*, HMS/MGH

ISAAC KOHANE, *Associate Professor of Pediatrics, HMS, Director, Children's Hospital Informatics Program, Associate Professor of Medicine,* HMS/Children's Hospital

ROBERT LUE, *Executive Director of Undergraduate Education*, FAS (MCB)

CHARLES MARCUS, *Professor of Physics*, FAS

JONATHAN MCDOWELL, *Astronomer*, CfA

VAN MCGLASSON, *Department Manager, Computation Facility*, CfA

DOUGLAS MELTON, *Chair, Life Sciences Council, co-Director Stem Cell Institute*, FAS

DANIEL MORIARTY, *Assistant Provost & CIO HU-Provost & SPH*

VENKY NARAYANAMURTI, *John A. and Elizabeth S. Armstrong Professor of Engineering and Applied Sciences, Professor of Physics, Dean of the Division of Engineering and Applied Sciences and Dean of Physical Sciences*, FAS

RICHARD O'CONNELL, *Professor of Geophysic*s, FAS

HANSPETER PFISTER, *Associate Director and Senior Research Scientist*, Mitsubishi Electric Research Laboratories, Cambridge, MA

EITAN RUBIN, *Head of Bioinformatics Bauer Center for Genomics Research*, FAS

BRUCE ROSEN, *Director, Massachusetts General Hospital Martinos Center, Professor of Radiology, Harvard Medical School*, HMS/MGH

FREDERICK ROTH, *Department of Biological Chemistry and Molecular Pharmacology*, HMS

JAMES RICE, *Professor of Engineering Sciences and Geophysics*, FAS/ DEAs & EPS

PHILLIP SADLER, *Director, Science Education Department*, CfA

MATT SCHNEPPS, *Director, Science Media Group*, CfA

BRIAN SEED, *Director of the Center for Computational and Integrative Biology at MGH/Professor of Genetics at HMS*, HMS/MGH

MARGO SELTZER, *Herchel Smith Professor of Computer Science, Associate Dean for Computer Science and Engineering in the Division of Engineering and Applied Sciences*, FAS

EUGENE SHAKHNOVICH, *Professor of Chemistry & Chemical Biology*, FAS

IRWIN SHAPIRO, *Director, Center for Astrophysics (outgoing),* CfA/FAS

JOHN SHAW, *Professor of Structural & Economic Geology*, FAS (EPS)

JOY SIRCAR, *Chief Information Officer*, FAS (DEAS)

KRZYSZTOF STANEK, *Assistant Professor of Astronomy*, FAS

CHRISTOPHER STUBBS, *Professor of Physics & Astronomy*, FAS

REBECCA WARD, *Lecturer on Systems Biology, Director of Research Affairs*, HMS

LEE-JEN WEI, *Professor of Biostatistics*, HSPH

GEORGE WHITESIDES, *Woodford L. and Ann A. Flowers University Professor, Professor of Chemistry*, FAS

STEVE WOFSY, *Abbott Lawrence Rotch Professor of Atmospheric and Environmental Chemistry*, FAS

STEPHEN WONG, *Associate Professor of Radiology, Director of HCNR Center for Bioinformatics*, HMS/BWH