DESCRIPTION: State the application's broad, long-term objectives and specific aims, making reference to the health relatedness of the project. Describe concisely the research design and methods for achieving these goals. Avoid summaries of past accomplishments and the use of the first person. This abstract is meant to serve as a succinct and accurate description of the proposed work when separated from the application. If the application is funded, this description, as is, will become public information. Therefore, do not include proprietary/confidential information. **DO NOT EXCEED THE SPACE PROVIDED.**

The Informatics Core is part of the Consortium for the "Collaborative Initiative on Fetal Alcohol Spectrum Disorders" (CIFASD). The theme of this collaborative initiative is a cross-cultural assessment of "fetal alcohol spectrum disorder" (FASD). The CIFASD will coordinate basic, behavioral, and clinical investigators in a multidisciplinary research project to better inform approaches aimed at developing effective intervention and treatment approaches for FASD. It will involve the input and contributions from basic researchers, behavioral scientists, and clinical investigators with the willingness to utilize novel and cutting-edge techniques so as not to simply replicate previous or ongoing work, but rather to try and move it forward in a rigorous fashion.

The Informatics Core will develop and maintain the CIFASD Data Repository, which will be used to collect, maintain and distribute data generated by the various participants in the consortium. The Informatics Core will be responsible for working with the other consortium participants to define a Data Dictionary to be used in standardizing data collection, enabling the transfer of data to and from the CIFASD Data Repository, consulting on how to establish local data management systems, providing both software tools and consulting to consortium participants, and producing status reports about the progress of the various projects within the consortium.

The Informatics Core draws on a wealth of resources, experience, and expertise at Indiana University in information technology infrastructure and data management, The CIFASD Data Repository will be developed on Indiana University's state-of-the-art central supercomputing facilities, taking advantage of Indiana University's strong commitment to institutional computational resources. Resources that will be used to implement the CIFASD Data Repository include multiple supercomputers, an array of readily available database and statistical software, and a high-speed, secure, robust data archiving system capable of storing duplicate copies in multiple physical locations separated by more than fifty miles.

The Informatics Core will use these extraordinary computational resources to provide a single, highly secure location for consortium participants to obtain the cross-cultural data that will enable the CIFASD to meet its goals of developing novel techniques for intervention and treatment of FASD.

PERFORMANCE SITE(S) *(organization, city, state)*
Indiana University
University Information Technology Services
2711 E. 10th St.
Bloomington, IN 47408-2671

Indiana University School of Medicine
Department of Medical and Molecular Genetics
975 West Walnut Street
Indianapolis, IN 46202-5251

KEY PERSONNEL. See instructions. *Use continuation pages as needed* to provide the required information in the format shown below. Start with Principal Investigator. List all other key personnel in alphabetical order, last name first.

Name	Organization	Role on Project
Craig A. Stewart, Ph.D.	Indiana University, University Information Technology Services	Principal Investigator
Anurag Shankar, Ph.D.	Indiana University, University Information Technology Services	Associate Co-Leader
Andrew D. Arenson	Indiana University, University Information Technology Services	Lead Developer
Stephanie L. Burks	Indiana University, University Information Technology Services	Database Administrator
Nathan A. Cosgray	Indiana University School of Medicine	Database Manager
Leah L. Flury	Indiana University School of Medicine	Applied Statistician
E. Chris Garrison	Indiana University, University Information Technology Services	Storage Systems Administrator

Disclosure Permission Statement. Applicable to SBIR/STTR Only. See instructions. ☐ **Yes** ☒ **No**

Research Plan

A. Specific Aims

The aims of the Informatics Core are to provide for data storage, quality assurance, data access, and data analysis to enable the Consortium for the "Collaborative Initiative on Fetal Alcohol Spectrum Disorders" (CIFASD) to amass and properly analyze the large quantity of data that will be required to help understand this sometimes subtle condition. Specifically, the aims of the core are as follows:

A.1) *The Informatics Core will support data management and submission to the CIFASD Data Repository by each of the Consortium for CIFASD's Programs, Cores, and Pilot studies.* The Informatics Core will consult and collaborate with each of the other programs, cores, and pilot studies participating in the Consortium for CIFASD as regards data storage and submission.

> A.1.1) *The Informatics Core will work with all Consortium for CIFASD participating units to create and disseminate a common data dictionary for all data shared within the consortium.* This will include variable names, labels, and valid ranges.

> A.1.2) *The Informatics Core will advise and consult with each of the other Consortium for CIFASD participating units so that the data storage mechanisms within each participating unit facilitate data transmission to the CIFASD Data Repository. The Core will create a manual for the submission of such data.*

> A.1.3) *The Informatics Core will create systems to permit submission of data in an encrypted format via high-speed networks. Data will be verified (according to the parameters identified in the data dictionary) before being added into the CIFASD Data Repository.* Encryption standards for all data submissions will meet HIPAA standards. An expert statistician will oversee validity checks on data submissions. These steps will ensure the security and validity of all data stored within the CIFASD Data Repository.

A.2) *The Informatics Core will maintain data stores in a reliable, secure fashion that will enable easy retrieval from properly authorized researchers.*

> A.2.1) *Data will be maintained in an Internet-accessible, secure, database. They will be easily accessible to those properly authorized to retrieve such data, and data will be delivered in formats specified by the requestor.* Critical to the success of the Consortium for CIFASD will be a system that enables the sort of data aggregation and large-scale analyses described throughout the proposal, so that it will be possible to properly analyze and understand FASD. A secure, Internet-accessible database will enable investigators to quickly and easily retrieve data needed to perform analyses. Data will be provided in the format desired by the requestor. Data will be stored and delivered in an encrypted format. The nature of the consortium requires careful control over data access, as it may be the case that not all data stored in the CIFASD Data Repository will be accessible to all researchers. Mechanisms for controlling data access will be built into the CIFASD data repository that will assure compliance with IRB approvals and scholarly collaboration agreements, assuring that all data retrievals are authorized.

> A.2.2) *Backup copies of all data will be stored in a secure, robust, long-term storage system.* We will implement a data storage mechanism that is secure and HIPAA compliant [1] by use of Indiana University's existing Massive Data Storage System [2], based on the use of the High

Performance Storage System (HPSS) software [3]. This data storage system will be robust in the preservation of data over time (decades) and robust in the face of natural disasters. The HPSS software used at IU will automatically detect aging or failing tapes and will copy data from such tapes to new tapes without loss of information. Data will be replicated within IU's distributed Massive Data Storage System. One copy will be kept on the IU Bloomington campus, and one on the campus of Indiana University Purdue University Indianapolis (IUPUI), some 50 miles away, thus providing protection against data loss due to natural disasters. All data that are potentially identifiable will be stored in an encrypted format.

A.3) *The Informatics Core will provide Statistical Support for Experimental Design, and support for statistical software tools.*

A.3.1) *The Informatics Core will provide consulting on experimental design and analysis and ensure that the data are analyzed and scaled in a fashion that makes possible comparisons across the research Programs, Cores, and Pilot Studies.* The international scope of the Consortium for CIFASD and the plans outlined in the Administrative Core description will require statistical analyses and scaling to ensure comparability within the large and diverse suite of data sources used to investigate FASD.

A.3.2) *The Informatics Core will provide consulting on software tools and, where needed, access to computing systems with statistical software available.* The Center for Statistical and Mathematical Software at Indiana University [4] will provide consulting on use of statistical packages (e.g. SAS, SPSS, and, if requested, others). Any participants in the Consortium for CIFASD who do not have access to fast computers with statistical software installed will be provided computing accounts on Indiana University's IBM SP Supercomputer (one of the largest university-owned supercomputers in the US) [5,6]. SAS, SPSS, Matlab, and a wide variety of other statistical and mathematical software are available on this computer.

A.4) *The Informatics Core will support the Administrative Core in management of the Consortium for CIFASD.*

A.4.1) *The Infomatics Core will provide quarterly summaries of data held in the CIFASD Data Repository, including summaries of new data submitted during the quarter.*

A.4.2) *The Informatics Core will provide ad hoc reports for the Administrative Core on request.*

B. Background and Significance

As stated in the Administrative Core Proposal, "While an abundance of evidence exists on the outcomes following prenatal alcohol exposure, one of the hindrances to developing a full set of criteria for diagnosing FASD has been the lack of adequate numbers of subjects at any single research site. This has limited the generalizations that have been possible from any single research location, and the control of a number of potential intervening variables. Therefore, this consortium will integrate researchers from a number of sites, including several international locations, to share common protocols, so that a large number of individuals can be assessed using similar procedures. With such collaborations, it will be possible to answer questions, which could not be answered easily by data from any individual study site. Furthermore, replication between sites differing in population and environmental characteristics would provide converging evidence for specific patterns of effects resulting from prenatal alcohol exposure. "

The challenge for the Informatics Core is important and easily stated: create, provide, and support the data storage, management, and analysis facilities (including statistical analysis) that will enable the Consortium for CIFASD to integrate the data from a large number of sites, including international sites, so that the goals of the Consortium may be successfully met.

The broad geographic distribution of the investigators and investigations requires that the Internet be used as a means by which data are exchanged. However, use of the Internet for data exchange places strong requirements on protocols for data transmission, and the use of an Internet-accessible data repository require that the repository itself have impeccable security.

Indiana University's leadership in cyber security is relevant to its ability to successfully guard the security of the data to be stored for the Consortium for CIFASD. IU has dedicated significant resources to handling IT security on its campuses. Specifically, a central IT Security Office (ITSO) has been established and staffed adequately to handle the many IT security issues that arise every day [7]. In addition, an IT Policy Office (ITPO) is responsible for overall policy, including IT security policies. IU's information and technology security policies are comprehensive and those deemed non-sensitive can be accessed at the ITPO web site [8].

Indiana University also plays a role in national cyber security. Indiana University has established, through an agreement with the National Infrastructure Protection Center, the nation's first Research and Education Network Information Sharing Analysis Center (REN-ISAC), to ensure the security of the nation's research and higher education computer networks [9].

Indiana University is well known for excellence in computing support [10]. Indiana University is a national and international leader in assessing and assuring that services provided to computer users are regarded positively by the users and are delivered in a cost-effective fashion [11-13]. In an externally administered random sample survey of Indiana University faculty, staff, and students, survey results typically indicate that 95% or more of the university population is satisfied or more than satisfied with the services provided by University Information Technology Services [14].

Indiana University is a national and international leader in high-speed networks. Indiana University operates the Network Operations Center for Abilene, the US's fastest production network, as well as several international linkages to other continents, regions, and countries [15].

Indiana University has achieved a leadership position in high performance computing, especially in support of biomedical research [16, 17]. Besides the relevant high-performance and biomedical research projects that will be detailed in the following section, the PI of the Informatics Core was also recently the lead author of a position paper regarding future opportunities in biomedical research through the use of high performance computing [18]. The staff involved in the Informatics Core, then, all possess recognized expertise in the delivery of excellent services in support of biomedical research.

C. Preliminary Studies

The leadership and staff of the Informatics Core have been involved in a wide variety of biomedical research through the activities of University Information Technology Services and the Indiana Genomics Initiative (INGEN) Information Technology Core [19]. As such, the leadership and staff of the INGEN IT Core have been deeply involved in many academic and clinical research projects. In

many of these areas staff and facilities serve a supporting or consulting role (see [20] for a partial list). In other areas, such as computational phylogenetics and radiation oncology, staff have led biomedical research projects and been authors or coauthors of research publications [e.g. 21-23]. Leaders and staff of the Informatics Core have also led the creation or modification of open source software for computational phylogenetics, modeling of complex biochemical systems, and radiation transport [24].

In the following sections we outline in detail areas of previous work particularly relevant to the activities of the Informatics Core.

C.1) *Massive Data Storage Service (MDSS)*

The proper management of data for the Consortium for CIFASD will require storage of large quantities of data in secure, robust, and reliable massive data storage systems. Indiana University (IU) established in collaboration with IBM a Massive Data Storage System (MDSS) to meet just such needs in 1999, well ahead of academic institutions worldwide.

IU's Massive Data Storage System is based on use of a software product called High Performance Storage System (HPSS) [3]. HPSS is an international collaboration of US national weapons labs (such as Los Alamos, Livermore, Sandia), NASA, supercomputer centers, and large scientific labs (such as the Brookhaven National Lab in the U.S. and RIKEN in Japan). Because HPSS is the product of a consortium, Indiana University has access to the source code and is thus assured that this system may be operated indefinitely.

HPSS provides data security and integrity sufficient to meet the needs of US nuclear weapons laboratories and well beyond the requirement specified in the HIPPA standards. Also, since HPSS is designed for the indefinite storage of data, it has facilities for detecting old or malfunctioning tapes, and for copying data from such tapes to new tapes.

Indiana University and the leaders of the proposed Informatics Core worked with HPSS and IBM to create the first production instance of geographically distributed data movers within HPSS [2,25]. Leaders of the Informatics Core led the creation of the software enhancements to HPSS that made possible the distribution and automatic copying of data between physically isolated tape robots. Data under the control of HPSS are stored in two StorageTek 9310 tape robots. One is located at the Indiana University Bloomington campus and the other at the Indiana University Purdue University Indianapolis campus, roughly 50 miles away (see Figure 1). Data are copied in nearly real time between the two StorageTek tape robots, using the I-Light optical network (owned jointly by Indiana University and Purdue University) [26]. Due to the existence of dual copies of all data, no data loss will occur even in the event that one of the buildings housing a tape robot were to be destroyed due to natural disaster or terrorism. The equipment that made the distributed storage system possible was made available in part by funding provided by the Indiana Genomics Initiative (INGEN). While biomedical research is replicable, any particular biomedical datum is not. Data that are lost are gone forever, and thus completely reliable preservation of data is of critical importance in biomedical research.
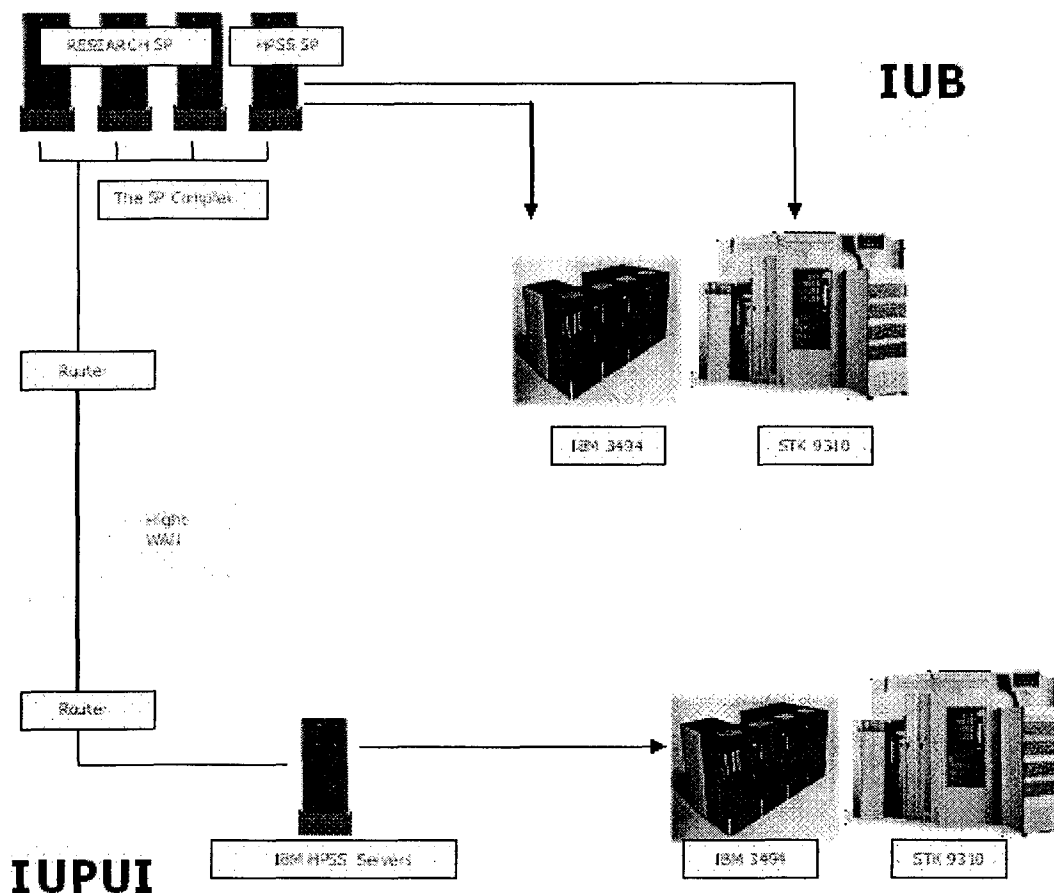
Figure 1. Indiana University's Geographically Distributed Massive Data Storage System

The total storage capacity of the Indiana University Massive Data Storage System is currently 500 Terabytes (TB; one TB is a million Gigabytes and holds roughly the same amount of data as 1,500 CDs). Indiana University stores more than a Terabyte of biomedical research data, one of the very largest (and perhaps the largest) such biomedical data store in the US.

C.2) *Centralized Life Sciences Data (CLSD) service*

The Centralized Life Sciences Data (CLSD) service provides a single, SQL-based interface to a selection of widely used biomedical datasets, including BIND [27], ENZYME [28], LIGAND [29], LocusLink [30], UniGene [31], dbSNP [32], SGD [33], KEGG PATHWAY [34] and a variety of NCBI BLAST databases [35]. CLSD runs on Indiana University's IBM SP Supercomputer and Sun E10000 supercomputer, using IBM DB2 database software and IBM's DiscoveryLink software [36] for retrieving data from multiple, heterogeneous data sources.

CLSD is already used by three research applications, including a Web interface designed to allow intuitive queries by biomedical researchers, a Web-based analysis platform for biomedical researchers who are customers of a DNA microarray facility, and as part of the biomedical data

reporting section of a laboratory information management system. Besides these production systems, there are multiple biomedical research applications in development that plan to make use of CLSD.

CLSD uses software programs called "parsers" to transform datasets from their native format into relational databases. Data files are copied across the Internet from their original sources to supercomputers at Indiana University. These files are converted to relational database format using a parser that is specific to each particular data source. The resulting relational database files are then imported into IBM's DB2 relational database management system. IBM's DiscoveryLink software and the CLSD software allow biomedical researchers to transparently query any or all of these diverse data sets without necessarily knowing where any particular piece of data originally resided. This allows, for example, the joining of data from different databases into a single record of information presented to the researcher. CLSD also incorporates BLAST datasets, permitting a researcher to run a BLAST job as if executing a database query.

Clinical and research labs can also make their data available to other IU investigators via CLSD. CLSD and its underlying database products enable clinicians and researchers to provide data both securely and with control over who has access to the data being provided. Data may be contributed via a wide variety of relational and non-relational formats, including DB2, Oracle, Sybase, SQL Server, Informix, flat files, Excel, and XML. This system allows a quick, simple, and secure means for sharing data while assuring that all researchers with rights to access the data are making use of the most current versions of those data. The CLSD system, user interface, and several of the underlying "'parser" programs were written by staff of the proposed Informatics Core.

C.3) *Data Repositories*

Members of the Informatics Core have also been involved in a number of collaborative studies in which Indiana University has been integral in data management. For example, investigators from the Department of Medical and Molecular Genetics at Indiana University led the National Cell Repository for Alzheimer Disease (NCRAD) [37]. This is a cooperative agreement (U24) with the National Institute of Aging, whose goal is to coordinate the recruitment of families with multiple members diagnosed with Alzheimer Disease. Recently, the scope of NCRAD has been expanded. It now also coordinates and databases the clinical information collected from 10 recruitment centers and maintains biological samples from all available family members. Subsequently, Alzheimer disease researchers from throughout the world can apply to NCRAD to obtain relevant samples or clinical data for their research. Thus, the database developed from this project has required extensive data quality verification and the ability to perform extensive data searches to identify samples meeting the researcher's specifications.

Researchers in the Informatics Core have also been critical to the success of the National Huntington Disease Research Roster [38]. Since 1979, this database of 2,380 families with Huntington Disease has been a resource of clinical data for researchers worldwide. In addition to obtaining detailed clinical information through the completion of questionnaires, the Roster also is a conduit for researchers seeking to identify study subjects for research. Thus, the database required for the Roster is extensive in the family history and clinical data that are stored, as well as in the detailed queries that must be completed to identify appropriate individuals who are then contacted to solicit their interest in a research project.

Another example of the extensive medical database expertise of the Informatics Core members is their ongoing role in the Collaborative Study of the Genetics of Alcoholism [39]. In its 14th year,

members of the Informatics Core, together with investigators at Washington University, St. Louis, have coordinated all data obtained from over 10,000 study subjects in 1,782 families, seen at any of the six recruitment centers. To ensure smooth operation in this large, collaborative study, monthly conference calls are held between the database managers at each of the six centers and the principal data management personnel located at Indiana University and Washington University, St. Louis. A secure Web-based interface was developed for this project and has proven to be an ideal mechanism by which study investigators and personnel evaluate data collection, generate reports, and query the database for critical data.

C.4) *Statistical Support for Experimental Designs*

Members of the Informatics Core are part of ongoing collaborative studies seeking to identify the genetic determinants of alcoholism and bipolar manic depressive illness. In this capacity, they participate in data validation and initial review of variable distributions, and provide recommendations for potential statistical analyses of data. Importantly, these analyses are often integrated across several analytic centers and require coordination to ensure their successful completion. The role of the statistical group has included detailed review of all variables collected and managed as part of the project, as well as development of data queries to identify outliers and incompatible data entries.

One of the members of the Informatics Core has already collaborated extensively with several members of the CIFASD researchers. This has resulted in the development of a database in which uniformly collected variables have been validated and maintained. The data collected from these different studies were then jointly analyzed to examine the predictive ability of cognitive performance to classify FAS and non-alcohol exposed children [40].

D. Research Design and Methods

D.1) *The Informatics Core will support data management and submission to the CIFASD Data Repository by each of the Consortium for CIFASD's Programs, Cores, and Pilot studies. The Informatics Core will consult and collaborate with each of the other programs, cores, and pilot studies participating in the Consortium for CIFASD as regards data storage and submission.*

> D.1.1) *The Informatics Core will work with all Consortium for CIFASD participating units to create and disseminate a common data dictionary for all data shared within the consortium. This will include variable names, labels, and valid ranges.*

>> D.1.1.1) Creation of a Data Dictionary: One of the primary goals of the Consortium is to take advantage of diverse populations around the world to collect the large-scale dataset required to properly address the challenges posed by FASD. To make use of these diverse populations, however, requires consistency in the way that data are generated and labeled. We will work with the other Consortium participants to understand the data they are collecting and, in, particular, to make use of the standardizations being created by the other cores for how to deliver and record batteries of tests and other experiments. We will create a data dictionary that defines how data should be represented, what labels mean, and what ranges of values are valid. This data dictionary will be distributed to all Consortium for CIFASD participants, and the Informatics Core will conduct training classes on use of the Data Dictionary (and other informatics facilities) via video conferencing or in person at the annual meeting of the Consortium to be held most likely at the Research Society for Alcoholism meeting.

D.1.2) *The Informatics Core will advise and consult with each of the other Consortium for CIFASD participating units so that the data storage mechanisms within each participating unit facilitate data transmission to the CIFASD Data Repository, and will create a manual for the submission of such data.*

D.1.2.1) Input formats: The Informatics Core will work with each of the other subunits of the Consortium for CIFASD so that each of these subunits can select a data management tool that will interoperate well with the Consortium efforts as a whole. The Informatics Core will not set particular standards, but an implicit requirement for each of the other participating units will be that they use some reasonably current data management tool (such as MySQL, Microsoft Access, SAS, or Excel). We will work with each participating subunit of the Consortium to create a process of selecting, formatting, and submitting data. We will most likely use Extensible Markup Language (XML) as an intermediate data exchange format. XML is a markup language that provides both a human-readable format for encoding and transmitting data and a method for defining what types of data are allowed to be in a file and how the data in that file relate to each other. XML will allow the use of standard tools for the creation, transmission, and verification of submissions across a wide variety of operating systems and programming languages.

D.1.3) *The Informatics Core will create systems to permit submission of data in an encrypted format via high-speed networks. Data will be verified (according to the parameters identified in the data dictionary) before being added into the CIFASD Data Repository.*

D.1.3.1) Data submission: The data submission process will use an intermediate staging area for submissions that are to be incorporated into the CIFASD Data Repository. Collaborators will be able to place submissions into the input staging area through a variety of methods, including both a Web page that allows files to be uploaded and direct access via secure FTP (SFTP). SFTP is a program that allows one to securely and quickly transfer files between computers. An automated process will periodically check the input staging area for new submissions, invoke data verification and quality checks on the submissions, and, if applicable, input the submissions into the CIFASD Data Repository.

Since the data submission process will provide valuable quality controls and feedback to our collaborators, it will be important for data to be submitted to the CIFASD Data Repository as promptly as is practical. Our collaborators, though, may need to have sole access to their data for an appropriate initial period. Collaborators will therefore be able to define the release date for their data and to specify the groups to which the data are to be released. Only data that have passed their release date will be made available to other researchers. The data submission process will include a check to make sure that the release date that's been attached to a set of results is valid.

D.1.3.2) Data quality checks and cleaning: Simple range checking programs will be utilized to screen for improbable and inconsistent data. These range checking programs can identify such errors as children with dates of birth occurring before their parents' dates of birth, individuals receiving evaluations on dates prior to their date of birth, as well as many other types of errors. These data checking programs will be

performed to ensure the accuracy and completeness of the data. A discrepancy database will be maintained, keeping track of potential errors and changes. The data from this database will be used to create all files for summary statistics, data exploration, and statistical analysis. Key fields will be indexed automatically and coded to enable fast and efficient searches through this database. Validation and integrity checks will be performed interactively to minimize data entry and reduce inconsistencies. Only specific answers will be allowed for each field. Data will be easily exportable to other software packages for statistical analysis, presentation, or correspondence. Users will be able to perform complex queries and generate sophisticated reports using a simple graphical user interface.

D.1.3.3) Data updates, relations, and revocations: Collaborators may need to submit data that directly update or are related to one or more previous submissions. The need to relate new submissions to existing data in the CIFASD Data Repository will be both incorporated in the schema for the CIFASD Data Repository and handled explicitly during data submission. The data submission process will use a mechanism whereby submitters can specify whether a submission represents new or updated data, and also indicates any other data to which the current submission relates. Most likely this control mechanism will be implemented using XML. Similarly, if data need to be removed from the CIFASD Data Repository, the appropriate XML can be formed to designate this action and specify which data are to be targeted for removal.

D.1.3.4) Intermediate processing: The CIFASD Data Repository will store not only initial results from the collaborating projects, but also secondary results created by processing the initial data. The Facial Imaging Core, for instance, intends to produce 3D images from the 2D images created at various locales. The Informatics Core will enable this intermediate processing by enabling the Facial Imaging Core to retrieve the 2D images and submit the newly created 3D images. The process used by the Facial Imaging Core to submit such processed images will be exactly the same as that used by any collaborator. The CIFASD Data Model will account for relating 3D images to the 2D images, and the Facial Imaging Core will then use the ability of the submission control mechanism to represent relations between the new 3D images and the existing 2D images in order to submit the 3D images appropriately. Other consortium members that want to process data and make the results available via the CIFASD Data Repository will use a similar process.

D.1.3.5) Privacy: Privacy will be maintained in strict compliance with HIPAA privacy regulations and IRB rules as well as with any further guidelines determined in conjunction with the Administrative Core. All files with any identifiable information will be stored in an encrypted format, and made viewable only to those people who have been properly authenticated and have the appropriate privileges to view the files. Appropriate security methods will be used to maintain these privacy policies.

D.1.3.6) Security: Security will be maintained in strict compliance with HIPAA security regulations. Secure Socket Layers (SSL), which enables data to be transferred in an encrypted format over the Internet, will be used for Web-based communication. Secure FTP (SFTP), which also enables encrypted data transfer, will be used for file transfer. TCP wrappers, software that provides an access control mechanism for incoming requests, will be used to specify exactly which computers can access the CIFASD Data

Repository. Appropriate security patches will be applied to all systems used to support the CIFASD Data Repository. Indiana University is currently building a firewall to protect its perimeter.

D.2) *The Informatics Core will maintain data stores in a reliable, secure fashion that will enable easy retrieval from properly authorized researchers.*

D.2.1) *Data will be maintained in an Internet-accessible, secure, database.* Critical to the success of the Consortium for CIFASD will be a system that enables the sort of data aggregation and large-scale analyses described throughout the proposal, so that it will be possible to properly analyze and understand FASD. A secure, Internet-accessible database will enable investigators to quickly and easily submit and retrieve data needed to perform analyses. The nature of the consortium requires careful control over data access to manage scientific and privacy concerns. Mechanisms for controlling data access will be built into the CIFASD data repository that will assure compliance with IRB approvals and scholarly collaboration agreements, assuring that all data retrievals are authorized.

D.2.1.1) Central Data storage and creation of a unified Schema: Each program will maintain its own full set of detailed data locally and only submit to the CIFASD Data Repository those data that are to be used in analyses that include data from multiple Programs, Cores, or Pilot studies. [Note: the Informatics Core will provide backup services for all data for any Program, Core, or Pilot Study that requests this, but this will be a service separate from the CIFASD Data Repository.]

In order to perform the sort of integrated analysis of data from multiple sites and multiple studies, it will be necessary to invest considerable effort into the creation of the unified CIFASD Data Repository. The steps in this process will be:

D.2.1.2) Requirements Analysis: We will conduct a survey of the collaborating projects to assess the current state of data collection and analysis and the expected future state of data collection and analysis with respect to the types and amounts of data being collected and the methods being used for data collection, storage, and analysis. The data survey will provide necessary information for the creation of a unified schema, will be used to determine which formats need to be supported for data input and output, and will indicate which projects will require help in creating tools for data collection, submission, and retrieval.

D.2.1.3) Data Distribution: We expect collaborators to engage in two types of analysis activities: Unstructured, where a researcher browses for simple results or interesting relationships, and structured, where well-defined, probably larger-scaled analyses occur. For more general, ad hoc querying and browsing we will provide a Web-based system for querying initial and intermediate results. As the project progresses, more specific and possibly more elaborate reporting and browsing interfaces will be developed to handle well-defined, possibly large-scale analyses as the need arises.

D.2.1.4) Output formats: For in-depth, statistical analysis, collaborators will require large amounts of well-specified data to be downloaded to their own systems. We will develop Web-based tools to output data in well-supported, standard formats like XML, SPSS and SAS as well as other data formats, as required. We will make use of the

DBMS/COPY [41] program where appropriate to create these output files. The DBMS/COPY program is capable of transferring data among more than 80 different formats. For ongoing, well-defined analysis projects at a collaborator's site, data will be automatically extracted and distributed.

D.2.1.5) Control over data access: For several reasons, it will be valuable for the study as a whole to have data contributed to the CIFASD Data Repository as promptly as possible. For example, the data submission process will provide valuable quality controls and feedback to Consortium participants, including fine tuning both the Data Dictionary and the data submission process itself. The CIFASD Data Repository will also be very useful to the Administrative Core in monitoring the overall progress of the project. Effective monitoring will require prompt submissions. Participants in the Consortium for CIFASD, however, may wish to limit access to particular portions of their dataset for an initial period of time proscribed by their collaborators or the Administrative Core. Thus, the CIFASD Data Repository will include mechanisms for specifying in fine detail which Consortium participants may have access to the data and when such access may take place.

D.2.1.6) Handling image files: Image files can be very large and are thus not stored directly within the database. The image files will still be stored online and encrypted, but the database will include information about the images and pointers to the images rather than the entire image itself. The Web interface to the CIFASD repository will incorporate the image files so that users of the CIFASD repository will be able to seamlessly retrieve image and non-image data. This may prove extremely valuable in trying to correlate dysmorphology data, either from 2D or 3D pictures with the neurobehavioral outcomes collected on the same individuals.

D.2.1.7) Data storage and processing requirements: The Consortium for CIFASD Data Repository will operate on a 7x24x365 basis, running on IU's 1 teraflops IBM SP supercomputer. (1 teraflops is a trillion mathematical operations per second. IU's IBM SP was ranked in November 2002 as the 112th fastest supercomputer by http://www.top500.org.)

IU's IBM SP is presently equipped with three terabytes of disk storage and more disk storage is added on an ongoing basis. (A terabyte is roughly equivalent to 1,500 CDs worth of data). Compared to even the very high end of our data storage estimates, one terabyte, we already have enough capacity to handle more than the entire project.

The Facial Imaging Core, for instance, which we expect to be one of the largest contributors of data, plans to use and create no more than 30 megabytes of data per subject for as many as 1,000 subjects. At most, this would use just one percent of our current capacity. In other words, the available storage capacity is well in excess of the expected needs of the Consortium for CIFASD studies.

D.2.2) *Backup copies of all data will be stored in a secure, robust, long-term storage system.* We will implement a data storage mechanism that is secure and HIPAA compliant [1] by use of Indiana University's existing Massive Data Storage System [2], based on the use of the High Performance Storage System (HPSS) software.

supercomputers in the US). SAS, SPSS, Matlab, and a wide variety of other statistical and mathematical software are available on this computer.

D.4) *The Informatics Core will support the Administrative Core in management of the Consortium for CIFASD.* In particular, the Informatics Core will:

D.4.1) *The Informatics Core will provide quarterly summaries of data held in the CIFASD Data Repository, including summaries of new data submitted during the quarter.*

Data summaries – indicating simply data types and numbers of records – will be created automatically within two weeks of the end of each quarter year. These reports will be submitted to the Administrative Core.

D.4.2) *The Informatics Core will provide ad hoc reports for the Administrative Core on request.*

As needed, the informatics Core will provide further statistics about the CIFASD Data Repository to the Administrative Core.

E. Human Subjects

E.1) *Protection of Human Subjects*

The Informatics Core will not be engaging in any data collection or any Human subjects interaction and will not unnecessarily store any identifiable data about a person. Some of the data to be stored, though, like facial images, are by their very nature identifiable.

As detailed in sections D.1.3.5, 'Privacy', and D.1.3.6, 'Security', the Informatics Core will maintain strict compliance with HIPAA privacy and security regulations, including storing data in encrypted formats, using secure methods for transferring and accessing data, and using appropriate authentication and authorization procedures.

The Informatics Core has submitted an application for review to Indiana University's Institutional Review Board, the Bloomington Campus Committee for the Protection of Human Subjects, which is currently pending approval. Individual certifications for all of the key personnel in the Informatics Core are included with this application.

E.2) *Inclusion of Women*

Human Subjects will be determined by other members of the CIFASD.

E.3) *Inclusion of Minorities*

Human Subjects will be determined by other members of the CIFASD.

E.4) *Inclusion of Children*

Human Subjects will be determined by other members of the CIFASD.

INDIANA UNIVERSITY

DOCUMENTATION OF EDUCATION
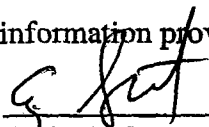## ON THE PROTECTION OF HUMAN SUBJECTS

As of October 1, 2000, the National Institutes of Health (NIH) requires all key personnel involved in the design or conduct of research involving human subjects to demonstrate that they have completed education on the protection of human research participants. While the education requirement may be fulfilled in a variety of ways and specific courses or reading materials are not prescribed, NIH does require that a description of the education completed be submitted prior to award. To document how you have fulfilled this requirement, please list below the education activities in the protection of human research subjects you have completed during the last five years and check "yes" if you have passed the Indiana University web-based certification test. Examples of appropriate educational activities include:

- **Readings**   *List books, brochures, or other materials concerning the use and protection of human research subjects which you have read and identify the author or source of the materials.*

- **Courses**   *List any courses, seminars, or workshops you have attended on the protection of human research subjects and the date you attended. Identify the organization or office that provided the training.*

- **Web-based Tutorial Programs**   *Identify any tutorial program you have successfully completed and the web site. For example, the NIH Office of Human Subjects Research training module located at http:/ohsr.od.nih.gov or the Indiana University training module located at http:/www.research.indiana.edu.*

- **Committee Participation**   *If you have served on an Institutional Review Board, please identify the committee, your role, and the dates you served.*

- **Other**   *List any other items or sources of training which have helped you fulfill the education requirement.*

I have taken and passed Indiana University's web-based Certification Test on the Protection of Human Subjects: ☒ **Yes**   ☐ **No**

I certify that the information provided below is correct.

Signature: _____   Date: 4/2/2003

Printed Name: Craig A. Stewart
Title: Director, Research and Academic Computing
E-mail address: stewart@indiana.edu

DESCRIPTION OF EDUCATION COMPLETED (include continuation pages if needed):

Dr. Stewart completed the web based Indiana University training module located at http://www.research.indiana.edu and certification test on March 24, 2003.

INDIANA UNIVERSITY

DOCUMENTATION OF EDUCATION
ON THE PROTECTION OF HUMAN SUBJECTS

As of October 1, 2000, the National Institutes of Health (NIH) requires all key personnel involved in the design or conduct of research involving human subjects to demonstrate that they have completed education on the protection of human research participants. While the education requirement may be fulfilled in a variety of ways and specific courses or reading materials are not prescribed, NIH does require that a description of the education completed be submitted prior to award. To document how you have fulfilled this requirement, please list below the education activities in the protection of human research subjects you have completed during the last five years and check "yes" if you have passed the Indiana University web-based certification test. Examples of appropriate educational activities include:

- **Readings**   *List books, brochures, or other materials concerning the use and protection of human research subjects which you have read and identify the author or source of the materials.*

- **Courses**   *List any courses, seminars, or workshops you have attended on the protection of human research subjects and the date you attended. Identify the organization or office that provided the training.*

- **Web-based Tutorial Programs**   *Identify any tutorial program you have successfully completed and the web site. For example, the NIH Office of Human Subjects Research training module located at http:/ohsr.od.nih.gov or the Indiana University training module located at http:/www.research.indiana.edu.*

- **Committee Participation**   *If you have served on an Institutional Review Board, please identify the committee, your role, and the dates you served.*

- **Other**   *List any other items or sources of training which have helped you fulfill the education requirement.*

I have taken and passed Indiana University's web-based Certification Test on the Protection of Human Subjects: ☒ **Yes**   ☐ **No**

I certify that the information provided below is correct.

Signature: _____   Date: 3 / 27 / 03

Printed Name: Anurag Shankar, Ph.D.
Title: Manager, Distributed Storage Services
E-mail address: ashankar@indiana.edu

DESCRIPTION OF EDUCATION COMPLETED (include continuation pages if needed):

Dr. Shankar completed the web based Indiana University training module located at http://www.research.indiana.edu and certification test on March 27, 2003.

INDIANA UNIVERSITY

DOCUMENTATION OF EDUCATION
ON THE PROTECTION OF HUMAN SUBJECTS

As of October 1, 2000, the National Institutes of Health (NIH) requires all key personnel involved in the design or conduct of research involving human subjects to demonstrate that they have completed education on the protection of human research participants. While the education requirement may be fulfilled in a variety of ways and specific courses or reading materials are not prescribed, NIH does require that a description of the education completed be submitted prior to award. To document how you have fulfilled this requirement, please list below the education activities in the protection of human research subjects you have completed during the last five years and check "yes" if you have passed the Indiana University web-based certification test. Examples of appropriate educational activities include:

- **Readings**   *List books, brochures, or other materials concerning the use and protection of human research subjects which you have read and identify the author or source of the materials.*

- **Courses**   *List any courses, seminars, or workshops you have attended on the protection of human research subjects and the date you attended. Identify the organization or office that provided the training.*

- **Web-based Tutorial Programs**   *Identify any tutorial program you have successfully completed and the web site. For example, the NIH Office of Human Subjects Research training module located at http:/ohsr.od.nih.gov or the Indiana University training module located at http:/www.research.indiana.edu.*

- **Committee Participation**   *If you have served on an Institutional Review Board, please identify the committee, your role, and the dates you served.*

- **Other**   *List any other items or sources of training which have helped you fulfill the education requirement.*

I have taken and passed Indiana University's web-based Certification Test on the Protection of Human Subjects: ☒ **Yes**   ☐ **No**

I certify that the information provided below is correct.

Signature: _Andrew D. Arenson_   Date: _3/31/2003_
Printed Name: Andrew D. Arenson
Title: Principal INGEN Data Specialist
E-mail address: aarenson@iupui.edu

DESCRIPTION OF EDUCATION COMPLETED (include continuation pages if needed):

Mr. Arenson completed the web based Indiana University training module located at http://www.research.indiana.edu and certification test on March 27,2003.

INDIANA UNIVERSITY

DOCUMENTATION OF EDUCATION
ON THE PROTECTION OF HUMAN SUBJECTS

As of October 1, 2000, the National Institutes of Health (NIH) requires all key personnel involved in the design or conduct of research involving human subjects to demonstrate that they have completed education on the protection of human research participants. While the education requirement may be fulfilled in a variety of ways and specific courses or reading materials are not prescribed, NIH does require that a description of the education completed be submitted prior to award. To document how you have fulfilled this requirement, please list below the education activities in the protection of human research subjects you have completed during the last five years and check "yes" if you have passed the Indiana University web–based certification test. Examples of appropriate educational activities include:

- **Readings** *List books, brochures, or other materials concerning the use and protection of human research subjects which you have read and identify the author or source of the materials.*

- **Courses** *List any courses, seminars, or workshops you have attended on the protection of human research subjects and the date you attended. Identify the organization or office that provided the training.*

- **Web–based Tutorial Programs** *Identify any tutorial program you have successfully completed and the web site. For example, the NIH Office of Human Subjects Research training module located at http:/ohsr.od.nih.gov or the Indiana University training module located at http:/www.research.indiana.edu.*

- **Committee Participation** *If you have served on an Institutional Review Board, please identify the committee, your role, and the dates you served.*

- **Other** *List any other items or sources of training which have helped you fulfill the education requirement.*

I have taken and passed Indiana University's web–based Certification Test on the Protection of Human Subjects: X Yes   No

I certify that the information provided below is correct.

Signature: *Stephanie Burks*   Date: 27 Mar 03
Printed Name: Stephanie Burks
Title: Principal Unix System Administrator
E–mail address: stlburks@indiana.edu

DESCRIPTION OF EDUCATION COMPLETED (include continuation pages if needed):

Ms. Burks completed the web based Indiana University training module located at http://www.research.indiana.edu and certification test on March 27, 2003.

INDIANA UNIVERSITY

DOCUMENTATION OF EDUCATION
ON THE PROTECTION OF HUMAN SUBJECTS

As of October 1, 2000, the National Institutes of Health (NIH) requires all key personnel involved in the design or conduct of research involving human subjects to demonstrate that they have completed education on the protection of human research participants. While the education requirement may be fulfilled in a variety of ways and specific courses or reading materials are not prescribed, NIH does require that a description of the education completed be submitted prior to award. To document how you have fulfilled this requirement, please list below the education activities in the protection of human research subjects you have completed during the last five years and check "yes" if you have passed the Indiana University web-based certification test. Examples of appropriate educational activities include:

- **Readings**   *List books, brochures, or other materials concerning the use and protection of human research subjects which you have read and identify the author or source of the materials.*

- **Courses**   *List any courses, seminars, or workshops you have attended on the protection of human research subjects and the date you attended. Identify the organization or office that provided the training.*

- **Web-based Tutorial Programs**   *Identify any tutorial program you have successfully completed and the web site. For example, the NIH Office of Human Subjects Research training module located at http:/ohsr.od.nih.gov or the Indiana University training module located at http:/www.research.indiana.edu.*

- **Committee Participation**   *If you have served on an Institutional Review Board, please identify the committee, your role, and the dates you served.*

- **Other**   *List any other items or sources of training which have helped you fulfill the education requirement.*

I have taken and passed Indiana University's web-based Certification Test on the Protection of Human Subjects: ☑ Yes   ☐ No

I certify that the information provided below is correct.

Signature: _____   Date: _5/21/01_

Printed Name: Nathan Cosgray

Title: Data Manager

E-mail address: ncosgray@iupui.edu

DESCRIPTION OF EDUCATION COMPLETED (include continuation pages if needed):

Mr. Cosgray completed the web based Indiana University training module located at **http://www.research.indiana.edu** and certification test on September 20, 2000.

INDIANA UNIVERSITY

DOCUMENTATION OF EDUCATION
ON THE PROTECTION OF HUMAN SUBJECTS

As of October 1, 2000, the National Institutes of Health (NIH) requires all key personnel involved in the design or conduct of research involving human subjects to demonstrate that they have completed education on the protection of human research participants. While the education requirement may be fulfilled in a variety of ways and specific courses or reading materials are not prescribed, NIH does require that a description of the education completed be submitted prior to award. To document how you have fulfilled this requirement, please list below the education activities in the protection of human research subjects you have completed during the last five years and check "yes" if you have passed the Indiana University web-based certification test. Examples of appropriate educational activities include:

- **Readings**   *List books, brochures, or other materials concerning the use and protection of human research subjects which you have read and identify the author or source of the materials.*

- **Courses**   *List any courses, seminars, or workshops you have attended on the protection of human research subjects and the date you attended. Identify the organization or office that provided the training.*

- **Web-based Tutorial Programs**   *Identify any tutorial program you have successfully completed and the web site. For example, the NIH Office of Human Subjects Research training module located at http:/ohsr.od.nih.gov or the Indiana University training module located at http:/www.research.indiana.edu.*

- **Committee Participation**   *If you have served on an Institutional Review Board, please identify the committee, your role, and the dates you served.*

- **Other**   *List any other items or sources of training which have helped you fulfill the education requirement.*

I have taken and passed Indiana University's web-based Certification Test on the Protection of Human Subjects: ☑ Yes   ☐ No

I certify that the information provided below is correct.

Signature: _____   Date: 2 4 may 2001

Printed Name: Leah Flury
Title: Applied Statistician
E-mail address: lflury@iupui.edu

DESCRIPTION OF EDUCATION COMPLETED (include continuation pages if needed):

Ms. Flury completed the web based Indiana University training module located at **http://www.research.indiana.edu** and certification test on September 28, 2000.

INDIANA UNIVERSITY

DOCUMENTATION OF EDUCATION
ON THE PROTECTION OF HUMAN SUBJECTS

As of October 1, 2000, the National Institutes of Health (NIH) requires all key personnel involved in the design or conduct of research involving human subjects to demonstrate that they have completed education on the protection of human research participants. While the education requirement may be fulfilled in a variety of ways and specific courses or reading materials are not prescribed, NIH does require that a description of the education completed be submitted prior to award. To document how you have fulfilled this requirement, please list below the education activities in the protection of human research subjects you have completed during the last five years and check "yes" if you have passed the Indiana University web-based certification test. Examples of appropriate educational activities include:

- **Readings** *List books, brochures, or other materials concerning the use and protection of human research subjects which you have read and identify the author or source of the materials.*

- **Courses** *List any courses, seminars, or workshops you have attended on the protection of human research subjects and the date you attended. Identify the organization or office that provided the training.*

- **Web-based Tutorial Programs** *Identify any tutorial program you have successfully completed and the web site. For example, the NIH Office of Human Subjects Research training module located at http:/ohsr.od.nih.gov or the Indiana University training module located at http:/www.research.indiana.edu.*

- **Committee Participation** *If you have served on an Institutional Review Board, please identify the committee, your role, and the dates you served.*

- **Other** *List any other items or sources of training which have helped you fulfill the education requirement.*

I have taken and passed Indiana University's web-based Certification Test on the Protection of Human Subjects: ☒ Yes ☐ No

I certify that the information provided below is correct.

Signature: _E. Chris Garrison_   Date: _3/31/2003_
Printed Name: E. Chris Garrison
Title: Principal Mass Storage Specialist
E-mail address: ecgarris@iupui.edu

DESCRIPTION OF EDUCATION COMPLETED (include continuation pages if needed):

Mr. Garrison completed the web based Indiana University training module located at http://www.research.indiana.edu and certification test on March 27,2003.

F. (n/a) Vertebrate Animals

G. Literature Cited

[1] The Health Insurance Portability and Accountability Act of 1996 (HIPAA). http://www.cms.hhs.gov/hipaa/

[2] Shankar, A. and Bernbom, G. 2002. Building a Massive, Distributed Storage Infrastructure at Indiana University. Proceedings of the 10th NASA Goddard Conference on Mass Storage Systems and Technologies, pages 285-289, College Park, MD, April 2002.

[3] High Performance Storage System. 2003. http://www.sdsc.edu/hpss/hpss1.html.

[4] Center for Statistical and Mathematical Computing. 2003. http://www.indiana.edu/~statmath/

[5] University Information Technology Services, Indiana University. Research Computing Systems and Facilities. 2003. http://www.indiana.edu/~rac/systems_facilities.html.

[6] University Information Technology Services. 2001. Indiana University and IBM Unveil The Nation's Largest University-Owned Supercomputer. http://www.indiana.edu/~uits/cpo/ibmsp/.

[7] Information Technology Security Office. 2003. http://www.itso.iu.edu/.

[8] Information Technology Policy Office. 2003. http://www.itpo.iu.edu/.

[9] Indiana University. 2003. IU joins in landmark cyberspace security agreement. http://newsinfo.iu.edu/news/page/normal/804.html.

[10] University Information Technology Services. 2003. What awards has the Knowledge Base won? http://kb.indiana.edu/data/aiwp.html.

[11] Samuel. J.V., Peebles, C.S., Noguchi, T., Stewart, C.A. 2002. "Gauging IT Support Strategies: User needs then and now." Proceedings of the SIGUCCS 2002 Conference, November 20-23, Providence, RI. Available from http://www.indiana.edu/~rac/staff_papers.html.

[12] Samuel. J.V., Wilhite, K.J., Stewart, C.A. 2002. "Getting More for Less: A Software Distribution Model." Proceedings of the EDUCAUSE Conference, October 1-4, Atlanta, GA. Available from http://www.indiana.edu/~rac/staff_papers.html.

[13] Peebles, C.S., C.A. Stewart, B.D. Voss, S.B. Workman. 2001. "Measuring quality, cost, and value of IT services in higher education." Proceedings of the American Quality Congress, Charlotte, NC, May 2001.

[14] UITS User Satisfaction Survey. 2003. http://www.indiana.edu/~uitssur/.

[15] University Information Technology Services. 2003. GlobalNOC (Global Research Networks Operations Center) at Indiana University. http://globalnoc.iu.edu/.

[16] Stewart, C.A., C.S. Peebles, M. Papakhian, J. Samuel, D. Hart, S. Simms. 2001. High Performance Computing: Delivering Valuable and Valued Services at Colleges and Universities. Proceedings of SIGUCCS, Portland, OR, October 2001. Available from http://www.indiana.edu/~rac/staff_papers.html.

[17] Stewart, C.A., and R. Repasky. 2003. High performance computing in support of university biomedical research: a successful implementation. Proceedings of BioITWorld Conference and Expo. Boston, MA. Available online from http://www.indiana.edu/~rac/staff_papers.html.

[18] Stewart, C.A., et al. 2003. Opportunities for Biomedical Research and the NIH through High Performance Computing and Data Management. Published by the Coalition for Advanced Scientific Computing (http://www.casc.org/paper.html).

[19] Indiana Genomics Initiative (INGEN). 2003. http://www.ingen.iu.edu/.

[20] University Information Technology Services. 2003. Publications, presentations, and artistic works stemming from use of UITS High Performance Computing Systems, Massive Data Storage Facilities, or Advanced Visualization Environments. http://www.indiana.edu/~rac/hpc/papers.html.

[21] Stewart, C.A., D. Hart, D. K. Berry, G. J. Olsen, E. Wernert, W. Fischer. 2001. Parallel implementation and performance of fastDNAml - a program for maximum likelihood phylogenetic inference." Proceedings of SC2001, Denver, CO, November 2001. http://www.sc2001.org/papers/pap.pap191.pdf

[22] R.B. Cruise and L.S. Papiez, Integral Equation Formulation of a Mixed Diffusion-Jump Model of Elastic Scattering. Nuclear Mathematical and Computational Science: A Century in Review, a Century Anew. American Nuclear Society, LaGrange Park, Illinois, 2003.

[23] R.B. Cruise, R.W. Shepard and V.P. Moskvin, Parallelization of the Penelope Monte Carlo Particle Transport Simulation Package. Nuclear Mathematical and Computational Science: A Century in Review, a Century Anew. American Nuclear Society, LaGrange Park, Illinois, 2003.

[24] RAC Online Resources. 2003. http://www.indiana.edu/~rac/online_resources.html

[25] Indiana University's Geographically Distributed HPSS. 2003. http://storage.iu.edu/geog.html.

[26] The Indiana Optical Fiber Initiative. 2003. http://www.i-light.iupui.edu/.

[27] Samuel Lunenfeld Research Institute University of Toronto. BIND: Biomolecular Interaction Network Database. 2003. http://www.bind.ca/.

[28] Swiss Institute of Bioinformatics. ENZYME: Enzyme Nomenclature Database. 2003. http://us.expasy.org/enzyme/.

[29] Bioinformatics Center Institute for Chemical Research Kyoto University. LIGAND: Database of Chemical Compounds and Reactions in Biological Pathways. 2003. http://www.genome.ad.jp/ligand/.

[30] National Center for Biotechnology Information. LocusLink. 2003. http://www.ncbi.nlm.nih.gov/LocusLink/.

[31] National Center for Biotechnology Information. UniGene. 2003. http://www.ncbi.nlm.nih.gov/UniGene/.

[32] National Center for Biotechnology Information. dbSNP. 2003. http://www.ncbi.nlm.nih.gov/SNP/.

[33] Department of Genetics at the School of Medicine, Stanford University. SGD: *Saccharomyces* Genome Database. 2003. http://genome-www.stanford.edu/Saccharomyces/.

[34] Bioinformatics Center Institute for Chemical Research Kyoto University. KEGG. 2003. http://www.genome.ad.jp/kegg/kegg1.html.

[35] National Center for Biotechnology Information. BLAST. 2003. http://www.ncbi.nlm.nih.gov/BLAST/.

[36] IBM, Inc. DiscoveryLink. 2003. http://www-3.ibm.com/solutions/lifesciences/solutions/discoverylink.html.

[37] Indiana University School of Medicine Department of Medical and Molecular Genetics. National Cell Repository for Alzheimer's. 2003. http://www.iupui.edu/~medgen/research/alz/alzheimer.html

[38] Indiana University School of Medicine Department of Medical and Molecular Genetics. National Huntington Disease Research Roster. 2003. http://www.iupui.edu/~medgen/research/hd/research_roster.html

[39] Washington University in St. Louis. COGA Collaborative Study on the Genetics of Alcoholism. 2003. http://zork.wustl.edu/niaaa/

[40] Flury L, Mattson SN, Kodituwakku PW, Adnams C, Turner G, Li T-K, Riley EP. Neuropsychological phenotypes for identifying children with fetal alcohol syndrome. Genetic Epidemiology, 23:264-314, 2002.

[41] DataFlux, a SAS company. dbPower DBMS/Copy. 2003. http://www.dataflux.com/data/conceptual/dbmscopy.pdf

H. (n/a) Consortium/Contractual Arrangements

I. Letters of Support

J. (n/a) Product Development Plan (SBIT/STTR phase II, Fast-track)