

# Implementation of a Distributed Architecture for Managing Collection and Dissemination of Data for Fetal Alcohol Spectrum Disorders Research

Andrew Arenson<sup>1</sup>, Ludmila Bakhireva<sup>2</sup>, Tina Chambers<sup>2</sup>, Christina Deximo<sup>1</sup>, Tatiana Foroud<sup>3</sup>, Joseph Jacobson<sup>4</sup>, Sandra Jacobson<sup>4</sup>, Kenneth Lyons Jones<sup>2</sup>, Sarah Mattson<sup>5</sup>, Philip May<sup>6</sup>, Elizabeth Moore<sup>7</sup>, Kimberly Ogle<sup>5</sup>, Edward Riley<sup>5</sup>, Luther Robinson<sup>8</sup>, Jeffrey Rogers<sup>1</sup>, Ann Streissguth<sup>9</sup>, Michel Tavares<sup>1</sup>, Joseph Urbanski<sup>3</sup>, Helen Yezerets<sup>1</sup>, Craig A. Stewart<sup>10</sup>

<sup>1</sup> Indiana University, University Information Technology Services, Indianapolis, IN 46202, USA. {aarendson, cdeximo, jlrogers, mtavares, yyezert}@indiana.edu

<sup>2</sup> University of California, San Diego, Department of Pediatrics, La Jolla, CA 92093, USA. {lbakhireva, klyons, chchambers}@ucsd.edu

<sup>3</sup> Indiana University School of Medicine, Department of Medical and Molecular Genetics, Indianapolis, IN 46202, USA. {tforoud, joaurban}@iupui.edu

<sup>4</sup> Wayne State University, Department of Psychiatry and Behavioral Neurosciences, Detroit, Michigan, USA. {joseph.jacobson, sandra.jacobson}@wayne.edu

<sup>5</sup> San Diego State University, Center for Behavioral Teratology, San Diego, CA 92120, USA. {smattson, kowens, eriley}@sdsu.edu

<sup>6</sup> University of New Mexico, Center on Alcoholism, Substance Abuse & Addictions, Albuquerque, NM 87106, USA. pmay@unm.edu

<sup>7</sup> St. Vincent's Hospital, Indianapolis, IN 46032, USA. ESMoore@stvincent.org

<sup>8</sup> State University of New York, Buffalo, New York 14260, USA lutherkr@buffalo.edu

<sup>9</sup> University of Washington Medical School, Department of Psychiatry and Behavioral Sciences, Fetal Alcohol and Drug Unit, Seattle, Washington 98195, USA. astreiss@u.washington.edu.

<sup>10</sup> Indiana University, Office of the Vice President for Information Technology, Bloomington, IN 47405, USA. stewart@iu.edu.

**Abstract.** We implemented a distributed system for management of data for an international collaboration studying Fetal Alcohol Spectrum Disorders (FASD). Subject privacy was protected, researchers without dependable Internet access were accommodated, and researchers' data were shared globally. Data dictionaries codified the nature of the data being integrated, data compliance was assured through multiple consistency checks, and recovery systems provided a secure, robust, persistent repository. The system enabled new types of science to be done, using distributed technologies that are expedient for current needs while taking useful steps towards integrating the system in a future grid-based cyberinfrastructure. The distributed architecture, verification steps, and data dictionaries suggest general strategies for researchers involved in collaborative studies, particularly where data must be de-identified before being shared. The system met both the collaboration's needs and the NIH Roadmap's goal of wide access to databases that are robust and adaptable to researchers' needs.

**Keywords:** Distributed Computing, Repository, Data Dictionary, Fetal Alcohol Spectrum Disorders.

## **1 Introduction**

This paper describes the information technology infrastructure required to support distributed research on uncommon diseases such as those represented by FASD – an infrastructure which allowed researchers to collaboratively create and use a common, pooled data resource to enable discoveries and insights that would otherwise not be possible. FASD refers to a range of debilitating effects on the central nervous system in children who were exposed to alcohol as a fetus as a result of maternal alcohol consumption [1]. A variety of FASD diagnoses, such as Fetal Alcohol Syndrome (FAS), Fetal Alcohol Effects (FAE), and Alcohol-Related Neurodevelopmental Disorder (ARND) are differentiated primarily by the presence or absence of facial features, but all share developmental abnormalities in adaptive functioning, attention and memory problems, distractability, learning problems, poor judgment, and fine and gross motor difficulties [2]. Fetal Alcohol Spectrum Disorders are entirely preventable through maternal abstinence; yet an estimated 1-2 out of 1000 children per year in the United States are born with FAS, the most severe form of FASD [3]. One of the roadblocks to research is obtaining significantly-sized populations to study. The research challenge presented by the relatively low incidence of FASD is magnified by difficulties in physical diagnosis as distinguishing features may depend upon race and age. Only a small number of experts are currently qualified to make a definitive diagnosis of FASD. The National Institute on Alcohol Abuse and Alcoholism (NIAAA, one of the National Institutes of Health) has funded an international research consortium called the Collaborative Initiative on Fetal Alcohol Spectrum Disorders (CIFASD) to accelerate research in FASD and in particular to create new ways to make authoritative, differential diagnoses of FASD. CIFASD has as one of its fundamental organizing principles that more rapid progress can be made in understanding and developing interventions for FASD by combining the efforts of multiple researchers at multiple sites.

### **1.1 Distributed Nature of CIFASD**

The Collaborative Initiative on Fetal Alcohol Spectrum Disorders consists of researchers and study populations from six countries. Some of the participating sites are in relatively remote areas with little access to the Internet. The privacy of the populations must also be protected, including careful management of subject identifying information and data which are inherently identifiable (e.g. facial images). The combination of worldwide collection of data and the need to protect subject privacy led to a data management strategy based on a distributed architecture that enabled local entry and management of data along with subsequent submission and sharing of those data via a central repository.

## **1.2 Commitment to Data Integration**

Because CIFASD includes research groups that had previously been operating independently, the consortium made a joint commitment to use common data definitions and data dictionaries in order to make it possible to integrate data from different sites. To enable secure and reliable storage of data and to make sharing and analysis of data across multiple sites and multiple research groups possible, CIFASD also agreed that shared consortium data would be stored in a central repository. A fundamental design feature agreed to by CIFASD was that data would be added to the central data repository on a record-by-record basis. A record typically consists of data from a single modality (e.g. test result, interview response, or image) for an individual child or parent. Records were allowed into the central repository only when fully in compliance with the data dictionary. Thus, another design constraint was building a mechanism that ensured that all records added to the central repository be unique and in compliance with the data dictionary.

## **1.3 Clinical Needs vs. Computer Science Research**

The tools required by the collaboration were built by the Informatics Core of CIFASD in close cooperation with members of projects and other cores from across the collaboration. There is an inherent conflict between meeting the immediate requirements of end users quickly versus allowing computer science collaborators to create innovative tools to meet those demands [4]. We dealt with this conflict by expediting an initial rollout of software using well-understood technology, while laying the groundwork to migrate the software in the future to use advanced cyberinfrastructure techniques.

# **2 System Design**

The practical requirements dictated by the nature of the consortium, distributed data collection, and data analysis objectives called for a two-tier system enabling local data entry (at times no connection to the Internet was possible), a central data repository, and a mechanism for verifying and uploading data to the central repository.

## **2.1 Data Dictionaries**

The creation of data dictionaries used across the consortium was critical to its overall success. Earlier attempts to analyze data across multiple research studies encountered problems due to the difficulties of integrating data from multiple sites. These problems included different definitions of terminology and different experimental designs that led to tantalizingly similar but not quite comparable datasets. The details of the data dictionaries and their content will be described elsewhere. Data dictionaries were created for modalities including dysmorphology, neurobehavior, prenatal alcohol consumption, 3D facial images, and control factors. For these

different sources of data the number of variables per modality ranged from 20 to greater than 800 variables. Data types included numeric, category, and text data as well as 2D and 3D images.

One of the most difficult challenges involved in creating data dictionaries for the collaboration involved reaching consensus on which measures to use for prenatal alcohol consumption. Issues included cultural differences between populations and an interest in providing high quality tools for future data capture while still incorporating data from pre-existing studies. Cultural differences include such things as: differences in interpretations as to the quantity of alcohol in a single drink or how much alcohol consumption is required for a 'binge'; differences in the types of alcohol consumed; and differences in the best ways to elicit accurate responses from interviewees. A committee of representatives from the projects in the collaboration were brought together to determine a core set of variables that could be generated from the various instruments used to capture prenatal alcohol consumption, including agreeing on definitions and valid ranges for these variables. The core data dictionary was designed for use by researchers with pre-existing data. At the same time an expanded data dictionary was created to encompass the tenfold larger set of variables that had been used by two of the collaboration's projects to capture the raw data required to calculate the core variables. This expanded data dictionary was designated as the standard for any new projects entering the collaboration. Because of the possibility that new projects might not be able to use the expanded data dictionary due to cultural differences, such projects would fall back to using only the core variables if necessary. In this way the widest range of research was supported, allowing comparisons of the smaller set of core variables across all projects while still allowing analysis of the wider range of raw variables for those projects which used the expanded alcohol and control data dictionary. Thus we avoided the problem noted in other fields such as microarray analysis [5] where competing standards prevent data from different sites from being integrated.

It is unrealistic to expect any requirements gathering process to anticipate all future needs of researchers. It has been shown, for instance, that combining coded fields with free text fields provides the fullest assessment of a clinical database [6]. Free text fields were provided to enhance the effectiveness of the data collected by the collaboration and later proved useful in understanding seeming discrepancies in the data that had been entered.

Similarly, the creation of data dictionaries proved to be an ongoing rather than onetime process. The data dictionaries' primary importance lay in forcing and formalizing agreement on precisely what data were captured so that the data could be integrated. The nature of the data being captured, however, changed over time. For the most part, required changes to the data dictionary involved extensions for new types of data being captured. At times, however, experience of users attempting to capture or analyze data led to uncovering differences in interpretations of the data dictionary that necessitated refinements or corrections. Further it is to be expected, as has been seen in other research [7], that clinical practice will change over time, requiring data dictionaries, software tools, and repositories to change as well.

One issue that arose on a regular basis was the need to clarify the allowable ranges for various measures. When a subject scored a value outside of the allowable range for a particular variable, the designated authorities from within the collaboration were

called upon to work with the data manager to determine whether the instrument was administered correctly or the allowable ranges needed to be expanded in the data dictionary.

## **2.2 Data Entry**

Individual projects designated data managers, responsible for working with the CIFASD Informatics Core to assure that data was submitted to the central repository. Data managers were sometimes, but not always, the same people who were responsible for data entry. Data entry for the text and numeric data was accomplished through the use of Microsoft Access® databases, running standalone on PCs. Microsoft Access® was selected because of the combination of the functionality it provides for data entry and its wide availability to researchers around the world, most of whom already had the software installed. Separate graphical user interfaces (GUIs) were created for each modality (Dysmorphology, Neurobehavior, and Alcohol & Control) that required manual data entry. See Fig. 1 for an example. The GUIs proved efficient for handling the common data entry tasks of providing drop-down menus for category choices, checking data types, and checking data ranges. The GUIs also helped solve less common problems as they arose, such as:

- Providing a series of tabs that allowed data managers to navigate quickly amongst different subsections of a dataset to be entered
- Caching metadata to be applied to some but not all of a series of subtests
- Providing a date entry tool that allowed data managers to choose the correct date despite regional differences in how dates are written in different parts of the world (e.g. Month/Day/Year in the US versus Day/Month/Year in Finland)

**Alcohol & Control Variables Data Entry Form**

Subject ID:

Date of Interview:   Examiner:

Location of Interview:  Language of Interview:

Ethic Category:  Country of Origin:

American Indian/Alaska Native     Asian     Native Hawaiian or Other Pacific Islander  
 Black or African American     White     Unknown or Not Reported  
 More Than One Race     Cape Coloured     Other Specify

1. How often do you have a drink containing alcohol?   
2. How many drinks containing alcohol do you have on a typical day when you are drinking?   
3a. How often did you have five or more drinks on one occasion?   
    Ever binge?   
3b. How often did you have four drinks on one occasion?   
3c. How often did you have three drinks on one occasion?   
3d. How often did you have one or two drinks on one occasion?:   
3e. How often in the past month did you have five or more drinks on one occasion?   
3f. How often in the past month did you have four drinks on one occasion?   
3g. How often in the past month did you have three drinks on one occasion?

**Fig. 1.** Screenshot of the beginning of one section of the data entry GUI for Alcohol & Control variables. This GUI entry tool holds over 250 unique variables which can be exported as XML and transferred to the CIFASD central repository. Note the use of tabs (*middle*) to allow data managers to quickly maneuver amongst subsections of the data to be entered.

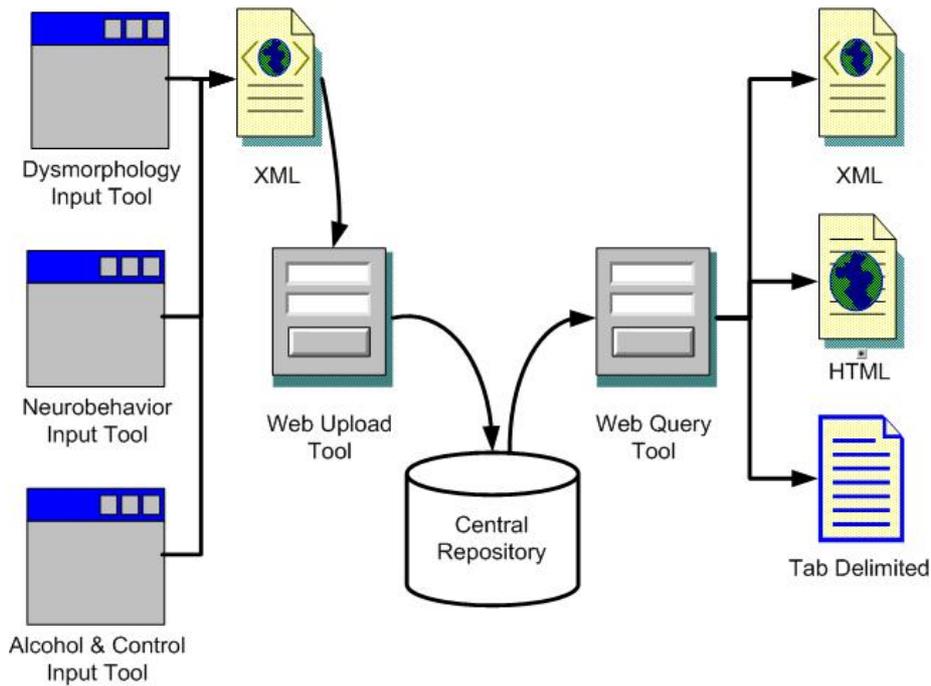
### 2.3 Central Repository and Data Flow

The central data repository consisted of an Oracle database with a Web-based front-end that accepted data submissions and allowed authorized users to search the database. Researchers could download data in a variety of formats suitable for browsing or importing into a statistical analysis program.

Data transfer to and from the central data repository was made available via XML. All uploads to the central data repository were done via XML, but downloads were also made available in other formats for ease of analysis in common statistics packages such as SAS and SPSS. The use of the standards-based XML format for data transport not only provided the necessary disintermediation for a scalable data

management architecture but also set the stage for a possible future migration to Web services and participation as part of a data grid.

The general architecture of the data flow diagram of the data management system for CIFASD is shown in Fig. 2.



**Fig. 2.** Diagram of data flow architecture for CIFASD. GUI-based data entry tools used to store data at local sites (*left*) produced XML-formatted, de-identified data, which was uploaded to the central repository and could be retrieved in a variety of formats (*right*).

The ability to export XML files was written into every local data entry tool, and a Web-based interface was created for uploading data in XML format to the central repository. The export of XML from the data entry tool at the local site was designed to exclude subject identifiers and rely instead on subject ID numbers that followed a globally-defined naming convention, ensuring subject confidentiality and enabling the ability to join data across modalities. Researchers were able to upload their local data via the Web-based submission tool and then receive one of four possible types of feedback, as follows:

- Data were compliant with the data dictionary, did not match any existing records, and were thus imported into the central data repository.
- Data were in some way incomplete or noncompliant with the data dictionaries and were thus excluded from entering the central repository.

- Data were an exact duplicate of existing data in the central repository, and thus no action needed to be taken, as the central repository was already up to date with respect to those data.
- Data appeared to be an update to an existing record. The data manager was then given an opportunity to review differences between the existing central repository version and the attempted upload and then queried as to whether or not the new record should have been considered an update of the prior record. New data identified as updates by the data manager were then applied to the central repository. Data not identified as a valid update by the data manager were ignored.

The distributed nature of the CIFASD architecture also allowed for flexibility in how different types of data were handled or in what tools local sites used for data entry. The 3D Facial Imaging Core, for instance, collected data that were generated from software and did not involve data entry. These data were still transferred to the central repository using a Web-based XML submission. Similarly, one of the local sites collected data using scanning technology and optical character recognition and did not require a separate data entry tool. Their data were accommodated by translating the results of their data collection process to XML before using the Web-based XML submission.

The data collected by the collaboration were complex in terms of representing a broad range of modalities: 2D and 3D facial images, brain images, facial examinations, neurobehavioral instruments, and questionnaires for prenatal alcohol consumption and control factors. In terms of the structure of the XML encoding and the central repository's relational database, much of this complexity was reduced through segregating by modality the variables and images that were to be collected and allowing data across modalities to be joined using the globally unique subject identifiers. For example, the dysmorphology data and the neurobehavior data had different concepts of uniqueness. Dysmorphology defined a unique record as being a unique tuple of subject id, examiner, and examination date while neurobehavior allowed a subject to have any subset of eighteen subexaminations but to not have any subexamination represented more than once. By grouping data first by modality, the peculiar needs of each modality could be handled individually within the structure of the XML file, within the code in the upload tool that parsed the XML and determined whether or not the data were compliant, and within the query tools that provided XML output to researchers.

Example of XML showing how complexity was handled via segregating variables by modality. Similar measures attained from 3D facial images and dysmorphological exams are combined in the same dataset. (Note: These data reflect that the same measurements were taken using different units in the two modalities – millimeters for the computer-calculated measurements from 3D facial imaging and centimeters for the hand-calculated measurements from dysmorphological exams. Measurement units and allowable ranges for all measures were strictly defined in the data dictionaries, but could vary between modalities.)

```
<Global ID="SMS43" >
```

```

<FacialImaging>
  <Scan Date="13-SEP-2005">
    <PhiltrumLength>14.77</PhiltrumLength>
    <RightPalpebral>28.02</RightPalpebral>
    <LeftPalpebral>28.06</LeftPalpebral>
    etc.
  </Scan Date>
</FacialImaging>
<Dysmorphology>
  <ExaminerLastName>Jones</ExaminerLastName>
  <DateExam>2005-08-16T00:00:00</DateExam>
  <PhiltrumLength>1.4</PhiltrumLength>
  <PFL>2.8</PFL>
  etc.
</Dysmorphology>
</Global ID>

```

## 2.4 Assuring Data Quality

The data dictionary's upfront purpose of forcing a shared understanding of the meaning of the data being collected was followed by the data dictionaries being used to ensure that only appropriate data were collected. This was implemented in three layers. The two primary layers were at the point of data entry and the point of data submission. The data entry tools included type and range checking as well as dependency checks between variables. The data submission tools repeated the type, range, and dependency checks both to accommodate the modalities and local sites that did not use a data entry tool, but also to safeguard against instances where the data could have been accidentally changed after data entry. The third layer of protection based on the data dictionary came from a check that can be run from within the data entry tool independent of data entry itself. Although unlikely, it was possible for researchers to bypass the type, range, and consistency checks during data entry, so the independent check allowed data managers to check their data locally to make sure that there were no inconsistencies before performing an export to XML and submission to the central repository.

## 3 Reliable Infrastructure

The data in the central repository were safeguarded via a variety of mechanisms. Access to the central data repository was controlled by authentication via a local Web-based authentication method – Indiana University's Central Authentication Service (CAS) [8]. This provided a ready way both to provision access to Indiana University Web servers by non-Indiana University personnel via its support for user-creatable Guest Accounts and to integrate with the Web-based applications used for data submission and retrieval. The servers that host the system were maintained by professional staff in modern computing facilities and were subject to regular security scans by a security group that operates independently of the system administration groups. The data resided on disk storage systems with built-in redundancy to protect

against the failure of any single disk. The data were backed up to the Indiana University Massive Data Storage System [9], which uses High Performance Storage System (HPSS) software [10] to write data simultaneously to automated tape libraries in two geographically distributed computing facilities (one in Indianapolis, IN; the other in Bloomington, IN). These automated tape libraries – and the Indiana University data centers in Indianapolis and Bloomington – are connected via the University-owned I-Light network [11]. This provided reliable storage of data and resilience against any sort of natural or manmade disaster short of a regional disaster affecting both data centers simultaneously.

#### **4 Results and Conclusions**

The data management facilities created for the Collaborative Initiative on Fetal Alcohol Spectrum Disorders proved effective in supporting the needs of the collaboration. The data integrity functionality embedded throughout the data management facilities provided the technical compliance assurance required for effective integration and analysis of the data collected by the members of the CIFASD. The distributed architecture of the system enabled researchers to collect data locally, maintain subject privacy, work in areas without robust Internet access, and still share data globally with the rest of the collaboration. Researchers were able to aggregate results from multiple populations for joint analysis in ways that were not previously possible and with a high degree of assurance in the quality of the data in the CIFASD central repository.

The critical proof of the data management infrastructure was in the quantity of subjects included and in the new types of science that were enabled. The CIFASD central repository accrued records for 224 subjects with FAS and as many as 463 that had at least some symptoms of FAS and might eventually be diagnosed as FASD. In comparison, one of the largest previous studies included only 88 FAS subjects [12]. Researchers were also able to analyze both cross-population datasets [13] and cross-modality datasets [14] on a scale that had never before been achieved. The increase in numbers of affected subjects provided a concomitant increase in the power of statistical analysis to detect differences amongst different subsets of the affected. The aggregation of results from many populations increased the range of possible hypotheses that could be examined, including comparisons amongst ethnicities and comparisons including some demographic and alcohol consumption categories that reached significant numbers only through inclusion of subjects from multiple sites.

The CIFASD data repository proved to be a practical, distributed system that leveraged current technology to provide new capabilities for important medical research. The most difficult operational challenges were in reaching consensus on the data dictionaries. The development of the CIFASD applications required some cleverness to handle the secure management of sensitive data, to manage data uploads and retrieval by a group of people with highly variable access to the Internet, and to resolve some issues that arose because the collaboration was international in scope (e.g. the matter of date formats). The use of XML for data transfer and the reliance on a commonly available commercial database product for data entry were key elements

in the rapid implementation of a system that successfully addressed the needs of the CIFASD researchers.

The total amount of data stored in the central data repository was relatively modest - roughly 2 GB. This should grow substantially over time. In the long run, we plan to migrate the CIFASD data management process and central repository to the NSF-funded TeraGrid [15]. One of the goals of the TeraGrid is to improve the general efficiency of scientific research within the US, and Indiana University is deeply involved in the TeraGrid as a Resource Provider. While the data management and analysis needs of the collaboration could easily be handled within Indiana University's facilities, using the infrastructure of the TeraGrid will enable better scalability and robustness of IU's services. Primary benefits will include leveraging the authentication and authorization infrastructure of the TeraGrid and increasing the disaster resilience of the CIFASD data repository. Copies of data are already stored in multiple locations within the Midwest region of the United States, but via the TeraGrid, it will be straightforward to also maintain backup copies of the data repository at locations that would remain safe even in the event of a regional disaster. Most importantly, leveraging the nationally-funded TeraGrid effort allows for economy of scale in the creation and management of cyberinfrastructure rather than crafting a new solution for each collaborative project.

The distributed data management architecture created in support of CIFASD served its purpose of supporting research and discovery related to the study of Fetal Alcohol Spectrum Disorders and is a useful model for any collaboration dealing with the needs of geographically separated sites that need to protect patients' privacy, deal with sometimes uncertain Internet access, and still allow researchers to share data. Further, the consortium created a persistent repository of interest that was and continues to be useful to studies of fetal alcohol spectrum disorders or related health impacts of maternal alcohol consumption. This persistent resource enabled more researchers to get at a wider set of important data and helped the consortium achieve the general NIH goal of "wide access to technologies, databases and other scientific resources that are more sensitive, more robust, and more easily adaptable to researchers' individual needs" [16]. The implementation of the CIFASD data management architecture was an effective use of distributed computing in support of a particular research project, which led to science results that would otherwise have been difficult or impossible to enable.

**Acknowledgments.** This research was supported in part by the Informatics Core for the Collaborative Initiative on Fetal Alcohol Spectrum Disorders, which received support from the National Institute on Alcohol Abuse and Alcoholism under grant 1U24AA014818. This research was also supported in part by the Indiana Genomics Initiative. The Indiana Genomics Initiative is supported in part by Lilly Endowment, Inc. The facilities that housed the CIFASD data repository were created with the assistance of grants-in-kind of hardware equipment from STK, Inc. (now a subsidiary of Sun Microsystems, Inc.) and Shared University Research grants from IBM, Inc. IU's participation in the TeraGrid is funded by National Science Foundation grant numbers 0338618, 0504075, and 0451237.

## References

1. Barr HM, Streissguth AP: Identifying maternal self-reported alcohol use associated with fetal alcohol spectrum disorders. *Alcohol Clin Exp Res*. 2001;25:283-7.
2. Streissguth AP, O'Malley K. Neuropsychiatric implications and long-term consequences of Fetal Alcohol Spectrum Disorders. *Semin Clin Neuropsych* 5:177-190 (2000).
3. Sampson PD, Streissguth AP, Bookstein FL, Little RE, Clarren SK, Dehaene P, et al. Incidence of fetal alcohol syndrome and prevalence of alcohol-related neurodevelopmental disorder. *Teratology*. 1997;56(5):317-26.
4. Hartswood M, Jirotko M, Slack R, Voss A, Lloyd S. Working IT out in e-Science: Experiences of requirements capture in a Healthgrid project. *Proceedings of Healthgrid*. 2005. IOS Press, ISBN 158603510X, Page 198-209.
5. Mattes WB, Pettit SD, Sansone S, Bushel PR, Waters MD. Database Development in Toxicogenomics: Issues and Efforts. *Environmental Health Perspectives*. 2004; 112: 495-505.
6. Stein HD, Nadkarni P, Erdos J, Miller PL. Exploring the Degree of Concordance of Coded and Textual Data in Answering Clinical Queries from a Clinical Data Repository. *Am Med Inform Assoc*. 2000;7:42-54.
7. Brindis RG, Fitzgerald S, Anderson HV, Shaw RE, Weintraub WS, Williams JF. The American College of Cardiology-National Cardiovascular Data Registry (ACC-NCDR): building a national clinical data repository. *J. Am. Coll. Cardiol*. 2001;37;2240-2245.
8. CAS basics. 2006. <http://uits.iu.edu/scripts/ose.cgi?akui.def.help>.
9. The Indiana University Massive Data Storage System Service. 2006. <http://storage.iu.edu/mdss.shtml>.
10. High Performance Storage System. 2006. <http://www.hpss-collaboration.org/hpss/index.jsp>.
11. Indiana's Optical Fiber Initiative – I-Light. 2006. <http://www.i-light.iupui.edu>.
12. May PA, Gossage JP, White-Country M, Goodhart K, Decoteau S, Trujillo PM, Kalberg WO, Viljoen DL, Hoyme HE. Alcohol Consumption and Other Maternal Risk Factors for Fetal Alcohol Syndrome among Three Distinct Samples of Women before, during, and after Pregnancy: The Risk Is Relative. *Am J Med Genet C Semin Med Genet* 127C:10-20 (2004).
13. Hoyme EH, Fagerlund Å, Ervalahti N, Loimu L, Autti-Rämö I. Fetal alcohol spectrum disorders in Finland: Clinical delineation of 77 older children and adolescents and comparison with affected North American children. Presented at the Western Society for Pediatric Research (February 2005).
14. Flury-Wetherill L, Foroud T, Rogers J, Moore E. The CIFASD Consortium, Fetal Alcohol Syndrome and 3-D facial imaging: Differences among three ethnic groups. Poster presentation at 29<sup>th</sup> Annual Research Society on Alcoholism Scientific Meeting, Baltimore, MD. (Jun 23-29, 2006).
15. TeraGrid. 2006. <http://teragrid.org>.
16. NIH Roadmap for medical research – Overview of the NIH Roadmap. 2006. <http://nihroadmap.nih.gov/overview.asp>