INDIANA UNIVERSITY

70b – IT Research – Life Sciences

Action Item Template Response

General Action Item Information

Lead Division/Office: Research Technologies
Action Item Number: 70b
Action Item Short Name: IT Research – Life Sciences
Dependencies with other EP Action Items: 3, 4
Implementation leader (name & email): William Barnett (barnettw@indiana.edu)

A. INTRODUCTION AND BACKGROUND

Important foundations in the development of a life sciences strategy for IT were laid at the *Empowering People* faculty retreat. Major needs presented were that Indiana University should:

- Continue its focus on information technology in the life sciences, and particularly the development of novel information technology and informatics approaches useful in the life sciences. Efforts in these areas should continue and be expanded, with particular focus on expansion of information technology useful in translational research and health science service delivery.
- Create mechanisms to support development of and interaction within the community of researchers who use advanced cyberinfrastructure at IU, such as some sort of center.
- Provide more support for data-intensive computing. Develop and deploy innovative hardware and software systems for data storage, and software tools for management of metadata, provenance, and data-centric workflows (addressed in EP Action Item 33b).
- Develop data mining, data processing/coding and data visualization tools for digital video and other multimedia data.
- Provide improved multimedia multi-streaming data mining, data visualization, and annotation. Driven by the ubiquitousness of multimedia data, this effort will support various video and audio data collected from multiple fields, from social and behavioral science departments, to music and library departments.
- Better support bioinformatics and computational biology, e.g., handling sequence data from the next generation sequencing technologies will be a good research topic for Informatics, Biology, Medical Sciences, and UITS.
- Provide trusted environments for the management and analysis of private (particularly HIPAA-implicated) data, which are key to the success of this enterprise and will raise the level of trust in UITS for all life sciences research.
- Provide high performance applications expertise so biological researchers can take advantages of high performance systems to advance their research.

In addition, Life Sciences has had conversations with Ruth Stone and IDAH about developing a community of software engineers so projects can more quickly ramp up by evening out the bumps in capacity, develop a cadre of skilled programmers who can be quickly deployed for projects, and promote the career growth of these programmers by providing training and shared experiences and collaboration.

Several possible current projects and project areas help identify more specific drivers for UITS life sciences efforts.

Area	Project	Value
Clinical Data Management	Institutional Clinical Data Repository (like i2b2, SHRINE)	Identifying opportunities for UITS to participate in national clinical research initiatives such as population studies, longitudinal research, patient recruitment, automated tissue collection, enabling more/better research, and sharing of clinical data among a number of sites.
Drug Discovery	3D protein model screening, Distributed Drug Discovery	Screening of 3D models of proteins to identify possible new therapies, improved health outcomes for neglected diseases, education for undergraduate and graduate students
Clinical Research Management	Case Report Forms	Enables researchers to manage clinical data on their own, through tools such as REDCap.
Web 2.0 Workflow	Tavern, Pegasus, Kepler, etc.	Developing workflow technologies that are easily usable by researchers and customizing those to support online analysis.
Genome Wide Association Studies	Personalized Medicine Research	Study of genetic variation across a genome used to identify genetic associations with phenotypes as a basis for personalized medicine approaches. Requires DNA sequencing and analysis of large populations.
Next Gen Sequencing	Supporting genomics technologies like SOLiD sequencers and microRNA studies	New sequencing equipment has the potential to generate vast amounts of data. The challenge is to develop methods to capture, manage and analyze these new data.
Cancer research	caBIG	New tools such as caBIG are changing the landscape for biomedical informatics and present new technology challenges.
Imaging	High throughput imaging	High throughput microscopes and other imaging technologies are creating new challenges for data acquisition, management, and analysis such as 3D object tracking.
Bioinformatics	Virtual Organization for Bioinformatics	Provide more easily accessed and reused expertise in finding, using, managing, sharing, and disseminating bioinformatics data.
Pharm-Tox Studies	National Gene Vector Biorepository, Multi-compartment model to predict drug exposure	Re-implementing the model algorithm using the C programming language and parallelizing the execution using OpenMP to take advantage of the Indiana University High Performance Computing systems and include more data points. Parallelization will permit

		rapid evaluation enhancements to the model and enable the performance of tests on multiple datasets quickly.
Image Processing	Cortical thickness changes in autism	Porting imaging algorithms to the IMB Cell B.E. architecture, an 8-way parallel processor. After tuning the implementation, the same analysis can be performed in less than a minute whereas before it had taken over 90 minutes.

I. DESCRIBE YOUR PLANS FOR IMPLEMENTING THIS ACTION.

Bioinformatics, proteomics, drug discovery, personalized medicine, population genetics (specific projects for Bioinformatics Staffing)

Aside from a few labs with high-throughput processes, the typical life sciences researcher is faced with the dilemma of needing specialized expertise in acquiring, handling, and analyzing bioinformatics data without the resources to hire their own staff to do so. Biologists have needs for sharing data, but are more concerned with how to manage and reuse their own data. We propose to provide a cyberinfrastructure and expertise to help individual researchers create, use, manage, analyze, and disseminate biological data. UITS would provide biomedical scientists and informaticians (bioinformatics workflow specialists and software engineers) to develop workflows so that biological and biomedical data are acquired with dissemination, provenance, and reuse in mind. We expect that researchers with tools to more easily acquire and analyze data will be more productive and more successful in making meaning of their data, disseminating results, and applying for funding. Over time, efficiencies would be found in the system as commonalities are discovered amongst different labs' needs. A virtual community could be built where researchers could gain access first to their own well organized data, but also get an idea of what other data researchers may have, and with whom they might then collaborate (this work undertaken in Action 4h - Virtual Organization Support). We propose to retain biomedical informatics scientists, computational biologists, high performance applications engineers for biological research, and software engineers who can understand the research workflows of medical researchers and biologists and create the tools to enable more efficient research efforts.

Metabolic Diseases & Cancer (specific projects for RTLS privacy efforts)

There is a rapidly growing need, and opportunity, to advance the research mission of the university through the management of private data for research. With the expanding partnership between UITS and the IU School of Medicine and the alignment of UITS systems with HIPAA regulations for securing electronic protected health information (ePHI), there is an opportunity to greatly increase the number of research projects with our partners in medical research and social science research. In addition, having committed to managing these data, UITS needs a sustainable risk management plan for RT systems, must be prepared to respond to incidents, and must position ourselves to accommodate new systems and services. The success, and national leadership, demonstrated in this effort requires additional resources if we are to continue to diligently secure ePHI and other sensitive or private data on our research systems. We propose to:

Continue investments in a strong and ongoing risk management plan for RT systems that includes required periodic review, modification of controls and risk postures in response to the changing regulatory landscape, and the provisioning of new, more researcher friendly, services demanded by our research partners. These will include the implementation of a public/private key server to allow researchers to more easily encrypt data and securely share it remotely with trusted parties.

Develop new tools and interfaces that enable easy, and more secure, access to RT services such as databases, data storage and archiving, private data sharing among multi-institutional teams, and access to HPC and other analytical resources.
Retain a Research Information Security Manager who will interpret the rapidly changing regulatory landscape, IU policy landscape, and the changing needs of researchers who need to work with sensitive or private information. This person will also guide RT and other UITS areas in the ongoing development of the Risk Management Plan for RT, required bi-annual reviews, incident management, and planning for the incorporation of new systems, software tools, and services as the research requirements for use of RT systems in the above described fashion continue to expand. This person will collaborate closely with RT systems managers, researchers, Information Assurance, and CACR for policy implementation.

Brain Sciences & Imaging Generally

The ability to visualize genomic and proteomic data, biological systems, health care IT surveillance, and population scale studies is becoming a critical distinguishing characteristic of evolving biomedical research, allowing scholars to effectively cope with and understand the exponentially increasing amount and variety of information that must be analyzed to move research programs forward. In partnership with the computational biologists and biomedical scientists, visualization specialists are required to enable life science researchers to better interpret their results. Many health care researchers, in particular, are MDs and come from clinical research backgrounds and so do not possess the technical backgrounds to master these technologies themselves.

Fueling the explosion of medical and biological imaging data are simultaneous increases in spatial resolution, temporal resolution, color/contrast depth, and affordability of various body imaging and microscopy devices. Researchers in the medical and life sciences have a number of common requirements derived from the need to store, organize, process, and analyze large collections of high-resolution images of 2, 3, or 4 dimensions. Many of these needs can be addressed through the acquisition and deployment of common software tools and provisioning of expert training and advanced support for those tools. Adoption of common tools will also facilitate greater sharing of data and expertise between research groups. Other aspects of the image processing and analysis pipeline would benefit from the extension of the university's cyberinfrastructure (CI) to address the unique needs of image-based research, including processing of ever-larger data sets, registration and integration of data from multiple imaging modalities, and interactive remote visualization and analysis tools.

In addition to these common needs, a number of IU researchers are not just applying the latest imaging technologies, but are conducting research into technologies and techniques that are helping to create the next generation of imaging systems. To sustain progress, these researchers have need for highly specialized expertise and deep support in areas such as automated image analysis, motion tracking, machine vision, and signal processing. Such deep collaborations between IT professionals and life sciences researchers will lead to proposals for external funding that will lead to long-term growth and leveraged sustainability for these collaborations.

In order to have a meaningful impact on the growing needs of the university's most strategic research community, we will need base funding for three individuals that will allow us to create an image processing, analysis, and visualization sub-group within UITS/RT with cross-campus responsibilities. These positions would include: (1) an imaging software specialist with a background in life sciences (B.S. or greater) and strong IT support and training skills, (2) an imaging CI specialist who merges understanding of various imaging modalities and needs with the ability to deploy and support tools in grid and cloud environments (M.S. or greater), and (3) one individual with deep expertise in image processing or machine vision (Ph.D. desirable) with the ability to develop new mathematical and algorithmic formalisms leading to breakthrough discoveries and transformational advances. Depending on their relative skill sets and levels of experience, one of (2) or (3) would also be tapped to be the team lead.

II. WHAT ARE THE POLICY AND PRACTICE IMPLICATIONS OF YOUR PLANS?

Research Technologies and UITS have committed to a trajectory of providing a trusted environment for the management and analysis of private data, particularly HIPAA-implicated data. Now that we are managing these data, IU should commit to their ongoing protection. In addition, new HIPAA regulations published by the FDA on January 16, 2009 and the prospect of yet more health care, and health care IT, legislation indicates a dynamic regulatory requirement, one that provides civil and criminal penalties for misuse of data. Indiana University should manage this changing risk landscape with continued diligence and investment in the secure management of our systems. This commitment opens the doors to participating in clinical research and providing medical school and other researchers who work with sensitive data (e.g.: social sciences research) access to a research cyberinfrastructure that puts them at a competitive advantage.

We propose that Research Technology software engineers form alliances with other IU entities (Information Assurance, OVPR and IDAH, other UITS areas) to create a more sustainable and flexible community in support of research projects.

I. III. IDENTIFY STAKEHOLDERS.

Life sciences researchers, particularly those involved in areas of research previously not well served by UITS, including clinical researchers at the IU School of Medicine, and those involved in new areas of biological and biomedical research.