- Develop tools in conjunction with early electronic health record (EHR) adopters to move the decision support to the physician-patient interface. These tools need to be managed by researchers to accomplish updates as the data from research is analyzed
- Continue to increase use and training in technology in medical school, residency, and post residency training through partnerships with educators and professional specialty organizations across all levels of healthcare providers
- Survey offices/facilities which are not adopting technology to identify reasons and what factors would improve rate of adoption. Human factors monitoring, some of which may be accomplished through remote network processes may be employed to identify bottlenecks, both user-originated and attributable to quality of service of software/network.
- Provide software research to assist physicians and facilities in identifying what is available and what the utility and efficiency factors are for different products
- Increase literature and seminar topics related to use of software with focus on patient outcomes
- Consider partnerships with insurance and/or drug detailers to provide computer basics to isolated practices.
- Increase data feedback of interest to physicians – personal, regional, specialty, etc, whatever data is available for analysis and incorporate information frameworks that address the totality of the health-providing environment, incorporating new and currently un-incorporated data components such as patient preference and consent. These items may create a demand pull if they also incorporate elements which satisfy the patient's ability and demand for information
- Encourage development of recursive adaptive information systems that evolve and accommodate user-specific knowledge structures
- Explore and make recommendations for how test-beds, model systems and the clinical trials infrastructure (e.g. the NIH roadmap, NECTAR, caBIG) can accommodate and incorporate the more recent advances in informatics software tools, using them to develop 'best practices' benchmarks

3.2 Breakout Group 2: Databanks for assessment of application-specific software and data-integration and other informatics tools.

Initial Charge to Panel: The ability to produce effective software tools for data analysis and clinical decision support is limited by the availability of the appropriate test datasets in reliable public databases. Effective datasets will require databases designed with shared data models and transparent interoperable data management. The goal of this session is to identify and assess the current status of available databases, to examine their underlying data models, and to define short and long term recommendations for the development of well-designed and interoperable data or database management systems

3.2.1 *Discussion Point:* *Identify the tools needed to facilitate the establishment of databases including tools for maintenance, for validation of data, and for effective and efficient usage (for both depositing data and accessing data). Develop recommendations for how the methods for image database collection can be better coordinated and integrated with clinical trials across NIH IC's, e.g., NIH Roadmap Re-Engineering the Clinical Research Enterprise/NECTAR, NIH Roadmap New Pathways to Discovery, NIH Roadmap National Centers for Biomedical Computing (NCBC), and efforts from the NCI Center for Bioinformatics.*

Specific recommendations for the national database resources include:

- Trans-NIH processes need to be developed to permit the collection of targeted data sets from on going clinical trails, and potentially linked to the current NIH NECTAR roadmap activity.
- The targeted data sets would include images, electrical signals, genomic and all other relevant patient data, together with truth information necessary to benchmark the performance of software tools.

- Security and quality control needs to be addressed as this database may serve as a standard for benchmarking software performance.
- The validation of input data and consensus process for the determination of truth files is critically important for the acceptance of performance standards.
- NIH support to maintain and update this resource is critically required.
- Leveraging of NIH support through public private partnerships is encouraged.

3.2.2 *Discussion Point*: *Identify database design requirements including the need for open-source tools to facilitate extraction of data from the database. Databases are not just for data storage – information retrieval vs. information extraction. Identify means for integration across databases.*

Recommendations:

- The development of repositories of representative imaging and clinical data is critical for evaluation and testing of new algorithms, as well as providing capabilities for integrating computerized analysis of existing and emerging imaging modalities and other data sources into clinical trials. There are two stages in the development:

- First Stage: Specific informatics tools required for data collection.
  - Ø Support development of the informatics tools and infrastructure
  - Ø Develop standards for user-friendly deposition of data by contributor (of primary data, metadata, and truth) and downloading data by others
  - Ø Implement rapidly using open source software development practices and grid technologies
  - Ø Should be applicable to multiple types of disease, images and other related data.
  - Ø Define inclusion/exclusion criteria for the data and redefine as necessary
  - Ø Sources of data can be from clinical trials or pre-existing individual lab databases.
  - Ø Require "certified" collection processes at each contributing institution to help ensure integrity of data, metadata, and truth.
  - Ø Should be started in parallel with the Second Stage.

- Second Stage: Actual design and establishment of databases -- Characteristics
  - Ø Each database needs an expert oversight committee and core support to check truth and to check unexpected results from users
  - Ø Database should be a repository, analogous to a tumor bank (i.e. continuously updated)
  - Ø Should be based on disease rather than imaging modality (include multi-modality and other data sources from different bio sensors)
  - Ø Should contain data and metadata including well-defined, multiple truths for the tasks that are related to the given disease, e.g., multimodality images of the breast for detection, diagnosis, treatment planning, response to therapy, etc.
  - Ø Determine database-specific minimal information standards for data submission similar to the MIAME standards for gene expression data.
  - Ø Include normal cases to serve as controls (e.g., these are the cases that may cause false positives)
  - Ø Include acquisition data (e.g., physical quality indices such as for image data)
  - Ø Include longitudinal data where appropriate (interval change)
  - Ø Include raw data as well as processed and/or reconstructed data for different biosensors
  - Ø Contributors and users need to commit to sharing
  - Ø Make incrementally available-- change database with changing technology
  - Ø Include ability to put derived data and results back into the database in an open source fashion

3.3.4 ***Discussion point:*** *Shared, Distributed Resources: identify the need for operation and distribution of resources for shared models, tools, and datasets, including research on their continued development. Types of shared and distributed resources, issues related to collaborative exploration environments and the role of government in building sustainable infrastructure were discussed. It was discussed that whether tools (i.e. software) with general purpose should be preferred to paradigm driven ones. In former case putting them into use for a specific need the issues of evolutionary development and a central pursuit were addressed. In general purpose tools, you need evolutionary and progressive annotated databases where data is tested and reported in the same central database. A good example is treating images as multivariate functions which could be used for visualization and diagnosis.*

Recommendations:

- Develop specifications for each of the stages so the collective force of the community can be unleashed.
- Be involved in making a national infrastructure for health maintenance and for sharing clinical data needs to be built and maintained.
- The operation and management of shared, distributed resources (community models, tools, datasets, computational servers), including ongoing development and enhancement activities.
- Consider the full life cycle of software tools and how to encourage development of a national infrastructure for sharing clinical (not just research) data.

3.3.5 ***Discussion Point:*** *Visual/Imaging-based Biomedical Informatics: identify the need for attracting research and development for visual/image-based data and integration into data mining development to include promotion of sophisticated image/data comparisons based on statistics, topology, semantics, and other abstract measures made of the non-textual data. Many advances still need to be made in fundamental computer science research, including multi-scale, multi-function visualization/imaging. Issues related to interactive access to multi-modal, multi-field data Error/uncertainty simulation and visual representation were discussed. Although, Imaging and visualization are ubiquitous, they are not coordinated. It's difficult for tools developers without access to representative groups of images. Data mining and data comparisons based on topology, statistics, model comparisons, and semantics are of interest and, therefore, integration of images with determining and network structuring would facilitate reaching these goals.*

Recommendations:

- Promote the visibility of image-based data and attract research and development for imaging, visualization, and image-based determining techniques and tools. Specific goals should include:
    - Ø promotion of sophisticated image/data comparisons based on statistics, topology, semantics, and other abstract measures made of the non-textual data
    - Ø Integration of image-based/non-text information with existing data mining methods.

- Invest in fundamental (not just applied) imaging and visualization research.
- Create inter-institute programs to develop general (as well as area-specific) software tools.
- Create an open database of images to use as "gold standards" for researchers.

3.3.6 ***Discussion Point:*** *Evaluate the Review Process for Development of Software and Tools: discuss the potential need for separate scientific and engineering evaluation groups, policies regarding evaluation of non-hypothesis driven research and development, and interdisciplinary team-based research.*

- Ø Participants in large, interdisciplinary projects
- Ø Leaders of distributed collaborations
- Ø Criteria reflecting key barriers to success
- Ø Evidence of previous interdisciplinary collaborations
- Ø Sound coordination/management plan

- Support more computer-science oriented study sections with focus on key NIH-relevant directions.

3.3.9 ***Discussion point:*** *Work flow: Tools will only have an impact on health if people use them, and they have to be within the context of people's workplace. One of important concepts should be that programming for health care and clinical decision making support cannot be traditional in the sense that it assumes regularity and determinism. Biomedical data is variable, ambiguous, and noisy: biological sequences, images, experimental observations, written descriptions. Problem-driven curriculum should be created. Rich domain provides motivation and problem sets for central areas of computer science (algorithms, databases, machine learning, and software engineering). Problem-driven courses need basic modeling and computational concepts learn relevant software tools, connect tools to create analysis/model, and manage computational resources (time, space, processors).*

Recommendations:

- Take on early deployment of a concept to demonstrate the usefulness of the concept and encourage industry to pick it up.
- Searching an archive for software development and issues is an incredible resource, and the NIH can assist with this so people can communicate with others in the field addressing the same software problems; this could spark collaboration.
- Consider funding specific programs to allow clinicians to build software. The clinician should specify what they want to do. The system should provide constant feedback about how far along it is. NIH could fund a study to interview clinicians and find out what it is that they want.
- Adopt a set of guidelines for specification to get the software engineers together on this topic. More high-level building blocks are important.