Overall description: Data Science and Applied Technology (DSAT) Core (RC4) provides many unique attributes, such as: developing software for interactive mobile technology (e.g., wearable sensors that are programmable in real time); validating new sensing technology; warehousing data; repurposing data; and applying machine learning techniques to domain problems. DSAT provides a central hub of expertise in computer science, biomedical engineering, biomedical informatics, data science, applied technology, epidemiology, and content expertise in the assessment of mobility.

Wearable sensor technology

- The core possesses 60 Actigraph GT1M, GT3X and LINK accelerometer models (The Actigraph Inc. Pensacola, FL). The monitors are small (3.8 x 3.7 x 1.8 cm), lightweight (27 g) and include a uniaxial and triaxial accelerometers. The accelerometers measure accelerations in the range of 0.05-2 G with a band-limited frequency of 0.25-2.5 Hz. The monitors are initialized, and data downloaded with the ActiLife software (Version 3.3.0).

- The core owns 15 Samsung Gear S smartwatches that possess customized software to program “apps”. Programming is done in TIZEN and Android operating systems. The applications loaded on the Gear S device run in a webkit based browser environment. Tizen provides API libraries to interface with its sensors as well as other system-level functionality and notifications.

- The core possesses multi-sensor technology through a portable armband (HealthWear Bodymedia, Pittsburgh, PA). The Sensewear armband uses a dual-axis accelerometer, a heat flux sensor, a galvanic skin response sensor, a skin temperature sensor, and a near-body ambient temperature sensor to capture data. Data from multi-sensor technologies are comparable to energy expenditure measured with doubly-labeled water.

- The core possesses 4 Empatica E4 wristband wrist worn wearable multi-sensor. The E4 measures blood volume pulse through a photoplethysmography Sensor - from which heart rate, heart rate variability (HRV), and other cardiovascular features are derived. An electrodermal Activity Sensor measures sympathetic nervous arousal and derives features related to stress, engagement and excitement. It also has a tri-axial accelerometer, event mark button and infrared thermopile for peripheral skin temperature.
Wearable technology validation and implementation services

- Validation against “gold-standard” measures of energy expenditure via indirect calorimetry and visual observation
- The core conducts focus groups and key informant interviewing to evaluate the acceptance of new technology. These groups help to optimize the adherence and retention in future studies utilizing this technology.
- The core possesses two k5 Cosmed Indirect calorimeters for validation of wearable technology

Mobility and activity measures using wearable technology

- Summary measures of physical activity include:
  - Total physical activity time (any type of activity at any intensity)
  - Time spent at specific intensities of physical activities
    - Sedentary
    - Light
    - Light-moderate
    - Moderate
    - Moderate-vigorous
    - Vigorous
- Mobility characterization
  - Total steps per day
  - Cadence (steps/min during active bout)
  - Step bouts – steps taken at a specific pre-determined cadence
- Basic GPS monitoring and tracking
  - Excursion size - average of maximum distance from the home for each excursion away from home
  - Excursion span - average daily maximum distance between all recorded locations away from home. Measures travel clusters, independent of maximal distance traveled.
- Geographical information systems for combining mobility patterns with the contextual environment
  - Geocoding and mapping according to CDC tracts
  - Adjacency/distance measures – for measuring distances between places and mobility patterns
  - Overlays – Points of interest (crime locations, walkways, sidewalks, parks, transportation services)

Wearable technology device administration, initialization and cloud computing

- A Map-Reduce framework (Apache Spark), as well as pre-defined scripts that leverage machine learning methods scaled for big data (SparkML), are included in the server software to be able to retrieve and analyze large amounts of data in real-time.
- Amazon web-services capability
- The server provides a platform to register participants, personalize the application based on their preferences, and configure data collection settings according to study requirements.
- Data collection configuration includes identifying active sensors, specifying their sampling rates, and defining the parameters used to aggregate the raw data into study-required variables.
- All of the configuration steps on the server are done remotely and without the requirement that the watches be collected from and returned to the participants.

Data repository

A data repository of de-identified data has been created for investigators to address age-related questions. A description of data available in the repository is listed in the table below.

<table>
<thead>
<tr>
<th>Study name</th>
<th>Sample size</th>
<th>Age (yrs)</th>
<th>Type of study</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIFE</td>
<td>1635</td>
<td>70+</td>
<td>RCT</td>
</tr>
<tr>
<td>LIFE-Pilot</td>
<td>424</td>
<td>70+</td>
<td>RCT</td>
</tr>
<tr>
<td>Women’s Health Initiative</td>
<td>161,808</td>
<td>50-79</td>
<td>RCT and obs. cohort</td>
</tr>
</tbody>
</table>

Section 20, Page 2
Big data analyses

- Time series data analyses and processing
- Supervised, semi-supervised and unsupervised machine learning
  - decision trees (random forest, classification of regression trees)
  - support vector machines, bag of words
  - deep learning
  - feature extraction
  - feature validation
  - pattern discovery
  - cluster analysis
  - models (support vector machines, artificial neural networks, deep learning)
- Epidemiological analyses [regression, random effects modeling, event time analyses (Poisson, Cox Hazard)]
- A suite of software to conduct a variety of analyses and visualization. These include STATA, SAS, SPSS, StatTransfer, MatLAB, Labview 16.0, R v3.4, JAVA and Enthought Canopy (Python).

Web portal development

- The portal provides information from all actively deployed watches and the collected data for each separate device.
- It presents summary statistics of activities, the current status of the watches, and detailed visualizations for the activity data.
- The data can be accessed through the web portal for data exploration and analysis

Data visualization

- Data are Visualized using GraphPad Prism 5.0, Tableau, Visual Studio, Adobe Photoshop, Adobe Illustrator and custom visualization packages in R.

Fast Healthcare Interoperability Resources (FHIR) SMART FHIR

- Fast Healthcare Interoperability Resources (FHIR) is a way for merging data formats for sharing electronic health records. It serves as a way to facilitate interoperation between mobile technology and electronic health systems. The core contains expertise for integrating mobile technology with FHIR exchange patient-generated data with the Epic® health record system and vice-versa.
- App Orchard for developing new EPIC interface apps

Consulting services. These services are provided due to the expertise as of the current investigators

- The UF Health Integrated Data Repository (IDR) and I2B2.
  - The core provides consulting services with regard to the IDR when used for aging-specific studies that meet the theme of the OAIC. The IDR enables new research discoveries as well as patient care quality and safety improvements through a continuous cycle of information flows between our clinical enterprise and research community. In its simplest form, a data repository is a collection of disparate data organized in a manner that lends itself to understanding relationships between data elements to answer questions.
  - The UF Health IDR currently consists of a Clinical Data Warehouse (CDW) that aggregates data from the various clinical and administrative information systems, including the Epic EMR. The CDW contains demographics, inpatient and outpatient clinical encounter data, diagnoses, procedures, lab results, medications, select nursing assessments, co-morbidity measures and
select perioperative anesthesia information system data.

- The CDW data contains “Fully Identified Data” and is fundamental to institutional business processes and secured per UF&Shands policies.
- Access to IDR data is provided through the NIH-funded i2b2 tool, which provides researchers access to a HIPAA-compliant and IRB-approved “Limited Data Set.” Faculty researchers can query the i2b2 Limited Data Set to identify cohort counts as they prepare grant proposals, plan clinical trials, and write IRB protocols.

**Partners**

- **E-health core.** Dr. Bian leads the e-Health core as part of the The University of Florida Health Cancer Center (UFHCC) Cancer Informatics Shared Resources program. This partnership, along with Dr. Bian investigator status on DSAT, allows for the two resources leverage knowledge gained in each program separately. The Shared Resources’ functions include:
  - Design, develop, and implement novel informatics methods, tools, and systems for capturing and synthesizing data to support clinical activities and clinical research;
  - Develop tools and methods to transform data collected by eHealth technologies, integrating with other relevant data sources, into actionable knowledge.
  - Develop and implement eHealth interventions and make eHealth tools freely available;
  - Support investigators to engage communities and key stakeholders in the development of eHealth tools and other patient- or clinician-facing technologies that are relevant to addressing the needs of patients, especially those in the UF Health catchment area, in collaboration with the UFHCC Community Outreach and Engagement (COE) Program;
  - Facilitate integration of tools into the electronic health record (EHR) systems and liaise with the EPIC/MyChart team at UF Health;
  - Liaise with the IDR/i2b2 team at UF Health when new data elements are being considered for i2b2; and
  - Liaise with UF Health IT security to ensure that tools being developed meet the security and privacy standards necessary.

- **OneFlorida Clinical Data Research Network.** RC4 actively utilizes the resources provided by the OneFlorida Data Trust. Following an investment of $100 million, in 2011 UF Health opened a new electronic medical record system and a clinical data warehouse that was the foundation for the development of an integrated data repository. Over the past 4 years, the IDR system expanded to the OneFlorida Network—a statewide Clinical Data Research Network (CDRN) that will join the PCORnet to optimize opportunities for conducting comparative effectiveness research (CER).
  - In 2012, One Florida cared for over 15 million unique patients, or about 70% of all Floridians, through a network of 22 hospitals, 1240 clinics/practices and 4100 providers.
  - The centerpiece of the One Florida CDRN is the OneFlorida Data Trust, a secure, de-identified data repository in which UF Health, Orlando Health, Florida Medicaid/CHIP, and the Florida Department of Health currently participate.
  - To date, the OneFlorida Data Trust houses data on 15 million patients, including demographic information, diagnoses, procedures, lab results, health care visit details, nurse assessments, bio-specimen availability, and vital statistics records.

- **intelligent HEAlth Lab (i-HEAL).** Dr. Rashidi’s laboratory is a close partner of DSAT. The lab focuses on: (1) transforming patient care in the Intensive Care Unit by developing autonomous monitoring tools using advanced machine learning techniques and (2) Developing intelligent tools for monitoring activity, cognitive and mental conditions of community-dwelling patients.
  - Resources and personnel are often shared between the two entities. The intelligent HEn mental Lab (i-HEAL) is located at the New Engineering Building (NEB). It includes desk space for up to 10 students. It contains a Dell Precision T5610 server with 64GB of memory and Dual Intel® Xeon processor and five networked workstations, each being equipped with four microprocessors. The software licensed to Dr. Rashidi’s lab for advance data analysis and programming include MTLAB, Visual Studio, Enthought Canopy (Python), WinEdt for LaTeX editing, and Microsoft suite for text and graphics processing.