CWRU SOM DATA REGISTRY PURPOSE

The CWRU SOM has established a Data Registry approved as exempt by CWRU IRB titled "Longitudinal Evaluation of CWRU School of Medicine's Curriculum." All data included in this registry are de-identified when used for educational research purposes.

The data in the registry are part of a Data Warehouse (DW) that enable faculty, staff, and students to:

1. Determine the extent to which the CWRU SOM fulfills its educational mission and reaches its goals.
2. Identify areas of curricular success and those requiring improvement.
3. Contribute broader understanding of teaching and learning in medicine.
4. Examine curriculum delivery in order to maintain quality standards and to ensure compliance with accreditation/licensure requirements.
5. Enhance understanding the effectiveness of teaching and methods that support learning.
6. Disseminate findings and lessons learned from CWRU SOM program evaluation activities to other medical education professionals through presentations and publications.

THE DATA REGISTRY CONSISTS OF:

1. De-identified, longitudinal database of learning, performance, quality assurance, and practice assessments of CWRU SOM students;
2. Student outcomes data and curriculum data

PRINCIPLES GOVERNING THE USE OF THE DATA REGISTRY:

1. **VALUABLE RESOURCE** that has value to the educational enterprise and is managed accordingly.

Rationale: Data resources are a valuable resource; it has real, measurable value. In simple terms, the purpose of data is to aid decision-making. Accurate, timely data is critical to accurate, timely decisions. Most assets are carefully managed, and data are no exception. Data resources are the foundation of our decision-making, so we must also carefully manage data to ensure that we know where it is, can rely upon its accuracy, and can obtain it when and where we need it.
Implications:

- Stewards must have the authority and means to manage the data for which they are accountable.
- We must make the cultural transition from "data ownership" thinking to "data stewardship" thinking.
- The role of data steward is critical because obsolete, incorrect, or inconsistent data could adversely affect decisions across the medical education enterprise.
- Part of the role of data steward, who manages the data, is to ensure data quality. Procedures must be developed and used to prevent and correct errors in the information and to improve those processes that produce flawed information. Data quality will need to be measured and steps taken to improve data quality - it is probable that policy and procedures will need to be developed for this as well.
- A forum with comprehensive institution-wide representation should decide on process changes suggested by the steward.

- Since data are an asset of value to the entire institution, data stewards accountable for properly managing the data must be assigned at the institution level.

2. **SHARED** Users have access to the data necessary to perform their duties; therefore, data are shared across institution functions and organizations.
   Rationale:
   Timely access to accurate data is essential to improving the quality and efficiency of institution decision-making. It is less costly to maintain timely, accurate data in a single application, and then share it, than it is to maintain duplicative data in multiple applications. The institution holds a wealth of data, but it is stored in many incompatible stovepipe databases. The speed of data collection, creation, transfer, and assimilation is driven by the ability of the organization to efficiently share these islands of data across the organization.
   Shared data will result in improved decisions since we will rely on fewer (ultimately one virtual) sources of more accurate and timely managed data for all of our decision-making. Electronically shared data will result in increased efficiency when existing data entities can be used, without re-keying, to create new entities.
   Implications:

- To enable data sharing we must develop and abide by a common set of policies, procedures, and standards governing data management and access for both the short and the long term.
- For the short term, to preserve our significant investment in legacy systems, we must invest in software capable of migrating legacy system data into a shared data environment.
- We will also need to develop standard data models, data elements, and other metadata that defines this shared environment and develop a repository system for storing this metadata to make it accessible.
- For the long term, as legacy systems are replaced, we must adopt and enforce common data access policies and guidelines for new application developers to ensure that data in new applications remains available to the shared environment and that data in the shared environment can continue to be used by the new applications.
- For both the short term and the long term we must adopt common methods and tools for creating, maintaining, and accessing the data shared across the institution.
- Data sharing will require a significant cultural change.
- This principle of data sharing will continually "bump up against" the principle of data security. Under no circumstances will the data sharing principle cause confidential data to be compromised.
- Data made available for sharing will have to be relied upon by all users to execute their respective tasks. This will ensure that only the most accurate and timely data is relied upon for decision-making. Shared data will become the institution-wide "virtual single source" of data.

3. **ACCESSIBLE (USER FRIENDLY)** to enable users to perform their functions.
   Rationale: Wide access to data leads to efficiency and effectiveness in decision-making, and affords timely response to information requests and service delivery. Using information must be considered from an institution perspective to allow access by a wide variety of users. Staff time is saved and consistency of data is improved.
   Implications:
   - Accessibility involves the ease with which users obtain information.
   - The way information is accessed and displayed must be sufficiently adaptable to meet a wide range of institution users and their corresponding methods of access.

- Access to data does not constitute understanding of the data. Personnel should take caution not to misinterpret information.
- Access to data does not necessarily grant the user access rights to modify or disclose the data. This will require an education process and a change in the organizational culture, which currently supports a belief in "ownership" of data by functional units.

4. **QUALITY** accountable for data quality.
   Rationale: One of the benefits of an architected environment is the ability to share data (e.g., text, video, sound, etc.) across the institution. As the degree of data sharing grows and curricular units rely upon common information, it becomes essential that only the data trustee makes decisions about the content of data. Since data can lose its integrity when it is entered multiple times, the data trustee will have sole responsibility for data entry which eliminates redundant human effort and data storage resources.
   Note:
   A trustee is different than a steward - a trustee is responsible for accuracy and currency of the data, while responsibilities of a steward may be broader and include data standardization and definition tasks.
   Implications:
    - Real trusteeship dissolves the data "ownership" issues and allows the data to be available to meet all users' needs. This implies that a cultural change from data "ownership" to data "trusteeship" may be required.
    - The data trustee will be responsible for meeting quality requirements levied upon the data for which the trustee is accountable.
    - It is essential that the trustee has the ability to provide user confidence in the data based upon attributes such as "data source".
    - It is essential to identify the true source of the data in order that the data authority can be assigned this trustee responsibility. This does not mean that classified sources will be revealed nor does it mean the source will be the trustee.
    - Information should be captured electronically once and immediately validated as close to the source as possible. Quality control measures must be implemented to ensure the integrity of the data.
    - As a result of sharing data across the institution, the trustee is accountable and responsible for the accuracy and currency of their designated data element(s) and, subsequently, must then recognize the importance of this trusteeship responsibility.

5. **CLEAR** Common Vocabulary and Data Definitions
   Data are defined consistently throughout the institution, and the definitions are understandable and available to all users.
   Rationale:
   The data that will be used in the development of applications must have a common definition throughout the Headquarters to enable sharing of data. A common vocabulary will facilitate communications and enable dialogue to be effective. In addition, it is required to interface systems and exchange data.
   Implications:
    - It is key to the success of efforts to improve the information environment. This is separate from but related to the issue of data element definition, which is addressed by a broad community - this is more like a common vocabulary and definition.

- ● Whenever a new data definition is required, the definition effort will be coordinated and reconciled with the corporate "glossary" of data descriptions. The institution data administrator will provide this coordination.
- ● Ambiguities resulting from multiple parochial definitions of data must give way to accepted institution-wide definitions and understanding.
- ● Multiple data standardization initiatives need to be co-ordinated.
- ● Functional data administration responsibilities must be assigned.

6. **SECURE** from unauthorized use and disclosure. In addition to the traditional aspects of national security classification, this includes, but is not limited to, protection of pre-decisional, sensitive, source selection-sensitive, and proprietary information.
   Rationale:
   Open sharing of information and the release of information via relevant legislation must be balanced against the need to restrict the availability of classified, proprietary, and sensitive information.
   Existing laws and regulations require the safeguarding of national security and the privacy of data, while permitting free and open access. Pre-decisional (work-in-progress, not yet authorized for release) information must be protected to avoid unwarranted speculation, misinterpretation, and inappropriate use.
   Implications:
   - Aggregation of data, both classified and not, will create a large target requiring review and declassification procedures to maintain appropriate control. Data owners and/or functional users must determine whether the aggregation results in an increased classification level. We will need appropriate policy and procedures to handle this review and declassification. Access to information based on a need-to-know policy will force regular reviews of the body of information.
   - The current practice of having separate systems to contain different classifications needs to be rethought. Is there a software solution to separating classified and unclassified data? The current hardware solution is unwieldy, inefficient, and costly. It is more expensive to manage unclassified data on a classified system. Currently, the only way to combine the two is to place the unclassified data on the classified system, where it must remain.
   - In order to adequately provide access to open information while maintaining secure information, security needs must be identified and developed at the data level, not the application level.
   - Data security safeguards can be put in place to restrict access to "view only", or "never see". Sensitivity labeling for access to pre-decisional, decisional, classified, sensitive, or proprietary information must be determined.
   - Security must be designed into data elements from the beginning; it cannot be added later. Systems, data, and technologies must be protected from unauthorized access and manipulation. Headquarters information must be safeguarded against inadvertent or unauthorized alteration, sabotage, disaster, or disclosure.
   - Need new policies on managing duration of protection for pre-decisional information and other works-in-progress, in consideration of content freshness.

EXAMPLES OF QUESTIONS THAT MAY INFORM CURRICULUM PLANNERS & DECISION MAKERS:

1. What is the relationship between NBME Customized Assessments administered at the end of each Block and SSEQ questions administered at the end of each Block?
2. Do students who achieve high scores on SSEQ exams (in Blocks 1 – 6) express different themes in their portfolios on professionalism than do students who have low scores on SSEQ exams (top 20 versus bottom 20)?
3. Do students who are identified for conscientious behaviors in two or more blocks express different themes in their portfolios on professionalism than students who have not been identified two or more times. (secondary question: what is the incidence of students who are identified as not meeting expectations on performing expected conscientious behaviors from 2006 - 2015)?