

September 2013

# Informatics Tools and Approaches To Facilitate the Use of Electronic Data for CER, PCOR, and QI: Resources Developed by the PROSPECT, DRN, and Enhanced Registry Projects

EDM Forum

AcademyHealth, [edmforum@academyhealth.org](mailto:edmforum@academyhealth.org)

Follow this and additional works at: [http://repository.academyhealth.org/edm\\_briefs](http://repository.academyhealth.org/edm_briefs)



Part of the [Health Information Technology Commons](#), and the [Health Services Research Commons](#)

---

## Recommended Citation

Forum, EDM, "Informatics Tools and Approaches To Facilitate the Use of Electronic Data for CER, PCOR, and QI: Resources Developed by the PROSPECT, DRN, and Enhanced Registry Projects" (2013). *Issue Briefs and Reports*. Paper 11. [http://repository.academyhealth.org/edm\\_briefs/11](http://repository.academyhealth.org/edm_briefs/11)

This Visualization is brought to you for free and open access by the EDM Forum Products and Events at EDM Forum Community. It has been accepted for inclusion in Issue Briefs and Reports by an authorized administrator of EDM Forum Community.

## Issue Brief

### Informatics Tools and Approaches to Facilitate the Use of Electronic Data for CER, PCOR, and QI

*Resources Developed by the PROSPECT, DRN, and Enhanced Registry Projects*

#### About the EDM Forum

The Electronic Data Methods (EDM) Forum is a three-year grant from the Agency for Healthcare Research and Quality (AHRQ) to advance the national dialogue on the use of electronic clinical data for the conduct of comparative effectiveness research (CER), patient-centered outcomes research (PCOR), and quality improvement (QI). The EDM Forum facilitates exchange and collaboration between eleven AHRQ-funded projects, including: the Prospective Outcome Systems using Patient-specific Electronic data to Compare Tests and Therapies (PROSPECT) studies; the Scalable Distributed Research Networks for CER, the Enhanced Registries for QI and CER, as well as other relevant healthIT initiatives. The EDM Forum and the research projects connected to the Forum are funded through the American Recovery and Reinvestment Act of 2009 (ARRA).

#### Executive Summary

Developing the health information technology (healthIT) infrastructure to support comparative effectiveness research (CER) was a core objective of the American Recovery and Reinvestment Act (ARRA) of 2009.<sup>1</sup> In response to substantial investments in this space—supporting more than 80 projects to build and test infrastructure using electronic clinical data (ECD)—innovative technology and strategies to support learning health systems have been developed that improve the ability to accurately capture, manage, and analyze ECD.<sup>2-3</sup> Recent contributions from clinical and biomedical informatics, as well as health services research, epidemiology, and clinical medicine have contributed to an array of tools and approaches for information management, care services, and research opportunities in the transforming health system.<sup>4</sup>

This brief describes 31 informatics tools and strategies that have been developed, adopted, or adapted by 11 of the ARRA-CER projects funded by the Agency for Healthcare Research and Quality (AHRQ) to improve the data infrastructure in health care and to demonstrate the ability to use ECD for CER, patient-centered outcomes research (PCOR), and quality improvement (QI). These efforts reflect the work of three grant programs: Prospective Outcome Systems using Patient-specific Electronic data to Compare tests and Therapies (PROSPECT), Scalable Distributed Research Networks, and Enhanced Registry projects. In this overview, the tools and approaches are grouped into four dimensions of data infrastructure that are necessary for network research or QI using ECD: 1) person-level data collection; 2) data access, exchange, and aggregation; 3) population level analytics; and, 4) provider, researcher, or patient-facing decision support.

The activities that are described are contributing a set of workable, extendable, and flexible informatics strategies, many of which can be adopted, adapted, and ideally, scaled for use in future research and QI activities. Descriptions of these informatics platforms and tools being implemented by the PROSPECT, DRN, and Enhanced Registry projects is provided to generate discussion about future applications. And while these 31 tools and approaches hold great promise to improve the transparency of research and QI, and accelerate diffusion of innovation in health care and outcomes, doing so will require continued support to achieve a culture shift towards collaborative science.

#### Introduction

Spurring innovation in health information technology (healthIT) infrastructure to support comparative effectiveness research (CER) was a core objective of the funding provided by the American Recovery and Reinvestment Act (ARRA) of 2009.<sup>5-6</sup> As a result of these investments there has been substantial development of new technologies and strategies to support the growth of learning health systems.<sup>7</sup> However, accurately capturing, managing, and analyzing data flowing through these systems in meaningful ways requires improving and expanding the existing data infrastructure (i.e., the data systems, research methods, governance policies, and professional training to do this work). In addition, an array of techniques for information management, care services, and research opportunities is transforming the ability of researchers to accelerate discovery and meaningfully use electronic clinical data (ECD) to improve patient outcomes.<sup>8</sup>



AcademyHealth

Of the ARRA-CER investments, \$100 million is managed by the Agency for Healthcare Research and Quality (AHRQ) to build the infrastructure for conducting CER with ECD. As part of this initiative, 11 research projects were funded through three programs: the Prospective Outcome Systems using Patient-specific Electronic data to Compare Tests and therapies (PROSPECT) studies; Scalable Distributed Research Networks (DRN) for CER; and, Enhanced Registries and QI and CER. This support reflects a forward-looking need to develop innovative ECD resources, and the need to facilitate use of these data for CER, PCOR, and QI. An additional AHRQ-funded project, the Electronic Data Methods (EDM) Forum, is directed by AcademyHealth to harmonize the efforts of the 11 research projects, engaging key stakeholders and facilitating synthesis and the dissemination of lessons learned in order to advance the science of research and QI using ECD. A background of these programs and profiles of the 12 projects is available in the EDM Forum's [CER Project Profiles Report](#).

All of the projects are addressing long-standing barriers to using ECD for research. For example, conducting CER across multiple sites is a challenge due to unintegrated systems, a lack of interoperability and standardization, and limited access to large heterogeneous datasets.<sup>9-10</sup> Many experts agree there is a critical need to improve the existing data infrastructure by developing new technology and innovative solutions to exchange, share, and access data for research and QI. A major charge of the PROSPECT, DRN, and Enhanced Registry projects has been to develop tools and informatics processes that facilitate exploration of new research questions, as well as the exchange and analysis of data across multiple institutions and settings.<sup>11-13</sup> The result of these efforts is a suite of workable, extendable, and flexible informatics strategies across the 11 research projects.

This brief describes 31 informatics platforms and tools that have been developed, adopted, or adapted by the PROSPECT, DRN, and Enhanced Registry projects as they apply to various dimensions in the development and execution of research and QI, including 1) person-level data collection; 2) data access, exchange, and aggregation; 3) population level analytics; and 4) provider, researcher, or patient-facing decision support.

A description of the major informatics platforms and tools being implemented by the projects within each of the four dimensions is provided, as well as a classification of the level of modularity of the resources to highlight the scalability to new research opportunities.

### Building the Electronic Clinical Data Infrastructure: The Role of HealthIT and Informatics for CER, PCOR, and QI

The informatics platforms and tools implemented by the PROSPECT, DRN, and Enhanced Registry projects support four dimensions of the infrastructure needed to process and analyze ECD for research and QI (**Figure 1**). The activities are provided to illustrate a generic model of the steps or functions required

#### PROSPECT Studies:

Comparative Outcomes Management with Electronic Data Technology (COMET)

Indiana PROSPECT

SURveillance, PREvention, and ManagEment of Diabetes Mellitus (SUPREME-DM)

Washington Heights/Inwood Informatics Infrastructure for Community-Centered Comparative Effectiveness Research (WICER)

The Population-Based Effectiveness in Asthma and Lung Disease (PEAL) Network

Enhancing Clinical Effectiveness Research with Natural Language Processing of EMR (CER Hub)

#### Scalable Distributed Research Networks:

SCALable National Network for Effectiveness Research (SCANNER)

Scalable Architecture for Federated Translational Inquiries Network (SAFTINet)

Scalable PARTnering Network for Comparative Effectiveness Research Across Lifespan, Conditions, and Settings (SPAN)

#### Enhanced Registries:

Surgical Care and Outcomes Assessment Program CER Translation Network (CERTAIN)

Building Modular Pediatric Chronic Disease Registries for QI and CE Research

#### Electronic Data Methods (EDM) Forum

to conduct CER, PCOR, and QI with ECD—particularly for networks that cross multiple institutions or care settings. The figure provides an evolving framework of the dimensions and subcomponents within to provide a high level illustration of the range of informatics strategies implemented to successfully use ECD for research and QI. These four dimensions include:

1. **Person-level data collection:** Tools designed to collect or gather data by integrating research and clinical care at the person-level, including data entry/collection at the point-of-care via EHRs; processing and abstraction tools to extract data and meaningful information from data sources, including natural language processing (NLP) of free-text from EHRs; and, devices or instruments designed for research data capture such as patient-reported outcome (PRO) tools.
2. **Data access, exchange, and aggregation:** Tools for accessing, integrating, and sharing data across and within institutions, including datamarts (i.e., the access layer of a data warehouse); data models; and, data grids.

3. **Population-level analytics:** Tools to support analysis at the population level to answer study questions, such as analysis of data quality or modeling, including query portals and explorers and distributed analytic tools.
4. **Provider, researcher, or patient-facing decision support:** Tools that leverage population analytics in patient care settings, including clinical and QI decision support; clinical research workflows; patient portals and navigators; and, mobile health (mHealth) applications.

A foundational component is the layer of governance and the fabric of trust needed to successfully allow the ‘flow of data’ while ensuring security and maintaining privacy. For further information on governance processes and policies supporting CER, PCOR, and QI please see the [EDM Forum eRepository Governance Toolkit](#).

It is important to note that the outlined dimensions do not necessarily need to be followed in chronological order, nor do they fully encompass all of the human and technical components in clinical care and QI, or those required to successfully complete a research study. In addition, there are standards, policies, and services that enable the data to flow from each component, as well as the need for validation of research results and formulation of clinical guidelines before evidence is disseminated and translated into clinical practice.

Each of the key dimensions of learning health systems that support research (data collection; access, exchange, and aggregation; analytics; and decision-support) are described in the sections below, including the informatics platforms and tools being implemented by the PROSPECT, DRN, and Enhanced Registry projects. Detailed descriptions of the informatics strategies are available in [Appendix A](#).

The table also indicates the level of modularity – or extent to which the tools can be adapted for new research and QI opportunities. These classifications are intended to inform further use, collaboration, and refinement of the tools. These distinctions are classified as follows:

1. **‘Plug-n-Play’:** Installable software which provide specifications that can be modified and customized to a new software environment, without the need for physical device configuration.<sup>14</sup> Generally, these tools do not rely on a specific technology or a vendor-based system (e.g., a specific EHR vendor).
2. **‘Hosted Platform’:** Tool or service hosted by and accessible within a network, often relying on a specific technology, data warehouse, or vendor-based system. Such services may be broadly accessible, or may require network membership.<sup>15</sup>

3. **‘Adaptable Approaches’:** Generalizable tools or approaches that can be borrowed or integrated into a new platform or environment, with additional programming or configuration to map to local systems.

Many of the informatics platforms and tools discussed in this brief are in relatively early stages of development and will require further validation and “hardening,” such as testing in additional informatics systems and architectures or testing applications with other diseases and outcomes. The purpose of this brief is to highlight the productivity thus far and inform future research opportunities. We encourage readers to explore the related resources further to find more information about the informatics platforms and tools and the research being conducted by the projects.

### Person-Level Data Collection

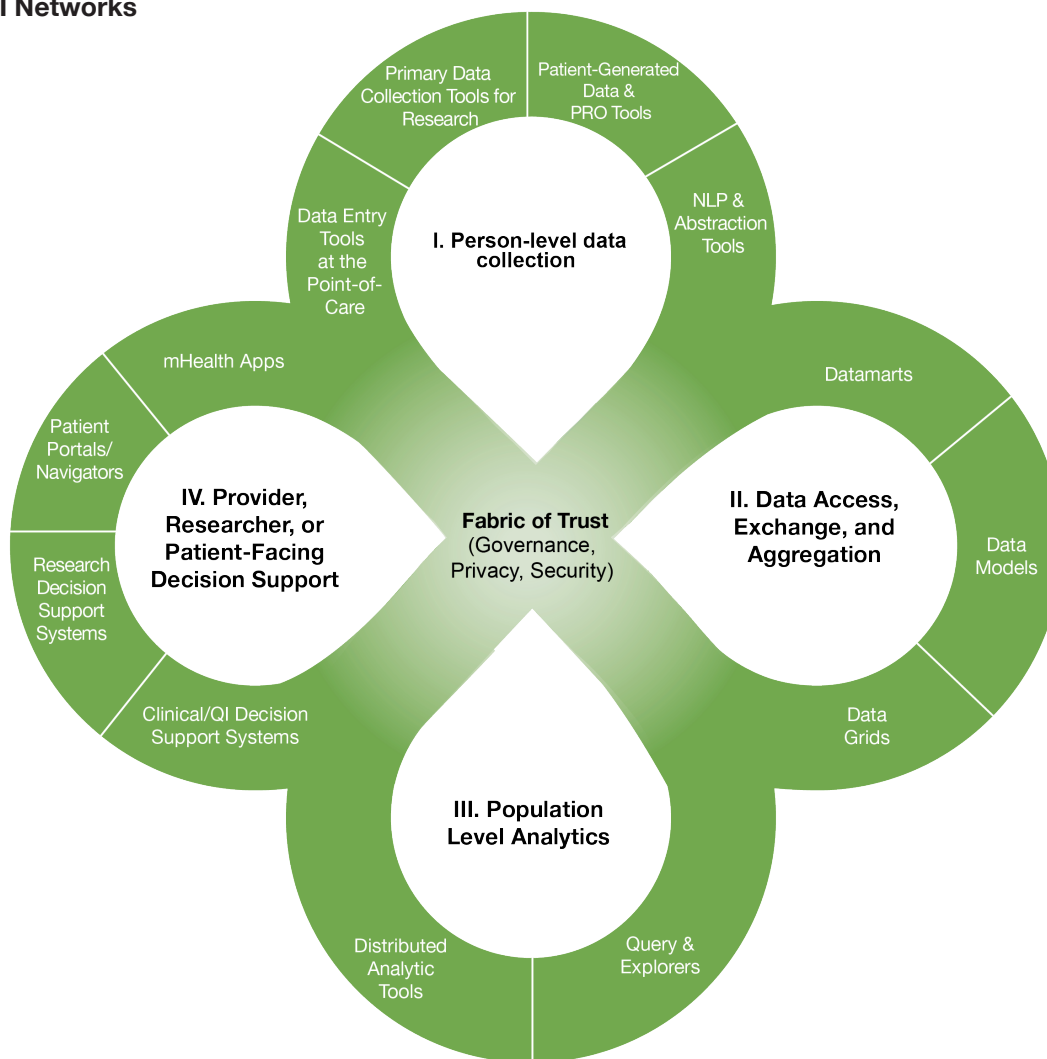
Data collection refers to the process of preparing and acquiring patient-level data both inside and outside the care process, for clinical, operational, and/or research purposes.<sup>16</sup> *Data collection* can occur at the point-of-care (e.g., via EHRs), in a closed research setting (e.g., monitoring blood pressure in a clinical trial), and through collection of PRO (see Table 1). The eight tools being developed for data collection improve the efficiency and quality of data while also expanding the scope and heterogeneity of population datasets. Some of these tools are integrated directly into the EHR to streamline the collection of research-grade data in the process of care. Other forms of research data collection are occurring outside the process of care, including innovations to meet the patient “where they are” by collecting patient-reported data and administering surveys using *mobile devices* such as tablets and smartphone apps.

Much of the clinical encounter is found in free-text (e.g. clinical notes) in the EHR. The difficulty of using this valuable information in research is that free-text lacks standardized vocabularies used to represent similar events across settings and contexts. These are useful to aggregate and conduct comparisons of events across settings. *Natural language processing (NLP)* identifies and codes isolated clinical concepts in order to extract meaningful information and represent events from free-text.<sup>17</sup> NLP and *abstraction tools* are used to extract this important information to supplement and gain insight on patients not available in structured elements.

### Data Access, Exchange, and Aggregation

Multi-site CER enables analysis of a larger and more heterogeneous population by leveraging data across sites and sources. In order to facilitate multi-site CER with ECD, processes and tools for data access, exchange, and aggregation are required (see Table 2). *Data access* refers to the authorization and processes of storing, retrieving, transferring or acting on data housed in a database, repository, or warehouse.<sup>18</sup> A *datamart*

**Figure 1: Informatics Tools and Approaches to Support the Use of Patient Level Electronic Clinical Data in Multi-Site Research and QI Networks**



is the access layer of the data warehouse environment that is used to get data out to the users, and are often designed around building cohorts for specific populations or conditions.<sup>19</sup> *Data exchange* is the process of combining, translating and integrating data structured in different ways and residing in different data sources.<sup>20</sup> Extract, Transform, Load processes, called ETLs, are included in these components. *Data grids* are technologies that facilitate the exchange of data, information, and knowledge resources to achieve interoperability across data sources. Finally, *data aggregation* is the process by which information is gathered across data held by different institutions and sources, and combined for comparisons and analysis.<sup>21</sup> Data aggregation is facilitated by *data models* which describe data elements, the way these elements are stored, and the relationship between elements in order to facilitate storage and retrieval of information.<sup>22-23</sup>

Linking individual patient data across data sources such as diagnostic (laboratory and imaging), pharmacy, claims (government and private), and EHRs is an important feature of many of the ARRA infrastructure projects. The ability to link patient information across

different care settings (inpatient, outpatient clinics, long term care, home care, emergency departments) is another important feature of record linkage activities. All of the PROSPECT, DRN, and Enhanced Registry projects are linking individual records across data sources and settings to some degree, and are facilitated by the nine tools presented in table 2 for data access, exchange, and aggregation. More information about the efforts to link data across the projects is provided in the CER Project Profiles Report.

**Population Level Analytics**

The four *analytic tools* that are presented support analysis at the population level to answer research or QI questions (see Table 3). *Query portals and explorers* facilitate the process of complex data queries, which are constructed to analyze diverse data elements, to identify important events, the sequence of these events, and their relationships to each other.<sup>24</sup> *Distributed analytics* facilitate the analysis across institutions’ datasets, often protecting security and privacy by distributing code or algorithms with no or minimal data transfer.<sup>25</sup>



Table I: Collection of Person-Level Data Tools				
	Title (Project)	Purpose and Function	Level of Modularity	Comments
Data Entry Tools at the Point-of-Care	*CareWeb Framework Clinical Plugins (Indiana PROSPECT)	An application framework that supports clinical plug-ins to be inserted into the process of care	Plug & Play	Open source version forthcoming
	*Enhanced eMR-ABC software (Indiana PROSPECT)	Using a web-based EHR, a software that supports the capture of clinical data, incorporates high volume results, and real-time trial enrollment technologies	Adaptable Approach	Uses the Aging Brain Care Mode; License required (see <a href="http://repository.academyhealth.org/egems/vol1/iss1/8">http://repository.academyhealth.org/egems/vol1/iss1/8</a> )
	*Open Source EHR-Based Registry (Pediatric Enhanced Registry)	A web-based data capture tool using i2b2 Forms Cell that supports entry of 'data-in-once' in center-specific databases. EHR data are extracted and upload via the EHR upload module to the registry	Adaptable Approach	Open source
Primary Data Collection Tools for Research/Patient-Generated Data & PRO Tools	Tablet App-Based and Web-based Data Collection (WICER)	A secure, HIPAA-compliant approach to collect and store primary data using tablet computers	Adaptable Approach	Approach can be adapted for other projects (see <a href="http://repository.academyhealth.org/egems/vol1/iss1/7">http://repository.academyhealth.org/egems/vol1/iss1/7</a> )
	*Patient Activation Tools: Inter-Visit Planner App (Pediatric Enhanced Registry)	Mobile app designed to assess the patient's symptoms and well-being in between clinic visits as well as provide a mechanism for parents to document any non-urgent questions	Hosted Platform	Proprietary mobile sensing platform. An account must be requested to use the platform for studies.
	Patient Voices Project Survey Center (CERTAIN)	Survey center services support data collection at the point-of-care, in a close research setting, or with PRO	Hosted Platform & Adaptable Approach	Contract services available to facilitate survey administration
NLP & Abstraction Tools	*Caradigm Amalga © (CERTAIN)	Middleware that semi-automates the abstraction and aggregation of EMR records for quality improvement	Hosted Platform & Adaptable Approach	Available within the suite of applications run on the Caradigm Intelligence Platform ©
	CER Hub Collaborative web-based informatics platform (CER Hub)	A platform to propose, develop, and conduct multi-institutional studies using MediClass application that apply NLP and structured data abstraction for EMR data	Hosted Platform	Membership in a CER Hub community is required. Projects must be approved by the CER Hub Steering Committee.

\* These tools are based on existing technologies or strategies that have been leveraged, expanded or improved by the projects (e.g., adopted or adapted for CER).

### Provider, Researcher, or Patient Facing Decision Support

The ten provider, researcher, or patient-facing decision support tools that are described leverage data or research findings to disseminate information back to clinicians or patients to inform health and health care decisions and to investigators to answer research questions. These tools can be integrated at the point of care, including clinical decision support, clinical research workflows, healthcare system patient portals, and mobile health (mHealth) apps.<sup>26</sup>

*Clinical decision support systems* are software used by clinicians or other health care providers to inform and provide the process of care for patients. Often, these systems are integrated with alerts to inform choices and knowledge based on information about the patient stored in the system or by data entered at the point-of-care.<sup>27</sup> *Research decision support* tools synthesize data to inform investigators about various research resources and

recommendations, for example alerts to prompt researchers to schedule visits or recruitment alerts for eligible patients. These tools help to improve the efficiency of conducting research in the clinical setting, whether for practical clinical trials or other clinical research designs. *Patient portals* are web-based applications for patients to interact with their own health care data and with their providers. Broadly speaking, *mHealth* tools are mobile communication devices to deliver health care services and provide decision support tools to patients.<sup>28</sup>

### Governance and the Fabric of Trust

Patient health data is characterized with proprietary and privacy restrictions, limiting the ability to retrieve and use ECD for research purposes.<sup>29</sup> In order to conduct rigorous research it is paramount to build trust among data sharing partners to facilitate the flow and use of data for analysis.<sup>30</sup> Getting buy-in from data partners can be achieved by demonstrating strong governance to

**Table 2: Data Access, Exchange, and Aggregation Tools**

	Title (Project)	Purpose and Function	Level of Modularity	Comments
Multi-Faceted Data Access, Exchange, and Aggregation Platforms and Tools	Data Cleaning Tool (Pediatric Enhanced Registry)	Converts raw data from the registry into a re-search-grade database along with its codebook	Adaptable Approach	Written in R
	*Diabetes DataLink (SUPREME-DM)	Multi-system distributed databases that capture and standardize detailed demographic and longitudinal ECD for diabetes surveillance	Hosted Platform	Uses the HMORN-VDW common data model
	Research Data Warehouse (WICER)	Research data warehouse provides de-identified data set of clinical data linked to WICER survey for participants who agreed to data linkage	Adaptable Approach	Approach builds on the New York-Presbyterian/Columbia University Medical Center data model
	Reusable OMOP and SAFTINet Interface Adaptor - ROSITA (SAFTINet)	A packaged data tool that transforms EHR and claims data to the OMOP common data model v 4.0, performs record linkage, runs data quality statistics, and removes personal identifiers. Data is then published to a grid-enabled database for use in distributed CER networks.	Plug & Play	Open source (see <a href="http://repository.academyhealth.org/egems/vol1/iss1/11">http://repository.academyhealth.org/egems/vol1/iss1/11</a> )
	Sleep Medicine Platform and Applications (COMET)	Platform supporting applications such as comparative effectiveness trials with instrument libraries, patient registries, portals, etc. Two major subsystems include 1) an open source data capture and processing pipeline built on an MS Windows OS and 2) a collaboration and presentation layer based on MS SharePoint.	Adaptable Approach	Currently focused on sleep medicine, but scalable in size and type to numerous applications
Datamarts	*PopMedNet Datamarts: Obesity and ADHD Modules (SPAN) and Asthma Module (PEAL)	Research user interface that permits menu-drive querying of multi-site data. These datamarts were developed within PopMedNet to access obesity, ADHD, and asthma data	Hosted Platform	Open source. Uses the HMORN-VDW common data model
	*HMORN Virtual Data Warehouse Common Data Model (SUPREME-DM, SPAN, CER Hub, PEAL Network)	A series of dataset standards and automated processes that allow SAS programs written at one HMORN site to be run against other VDW sites in a distributed fashion, without a central store of data	Adaptable Approach	Membership in the HMORN is required (see <a href="http://www.hmresearchnetwork.org/resources/tools/HMORN_VDW-Questions-and-Answers.pdf">http://www.hmresearchnetwork.org/resources/tools/HMORN_VDW-Questions-and-Answers.pdf</a> )
	*OMOP Common Data Model Version 4 (SAFTINet and SCANNER)	A common data model that enables researchers to standardize the format and vocabularies or coding schemes used across different data sources and types	Adaptable Approach	Open Source with minimal restrictions. Licensed under the Apache License Version 2.0
Data Grids	*Translational Informatics and Data Management Grid - TRIAD (SAFTINet)	Serves as the trusted communication and grid networking fabric	Plug & Play & Hosted Platform	The Biomedical Informatics Program at The Ohio State University (OSU) hosts the core TRIAD services and functionality on which SAFTINet relies. It is also possible for a network to implement and host TRIAD independently

\* These tools are based on existing technologies or strategies that have been leveraged, expanded or improved by the projects (e.g., adopted or adapted for CER).

protect privacy and security, and also by developing relationships among data sharing partners and demonstrating value of using the data for research. Some of the technical governance measures can be facilitated by the informatics platforms and tools, including policy models and guidelines integrated into data access tools to automate security and investigator approval processes. To guide the implementation of these processes, the SCANNER project has developed a **Privacy and Security Policy Framework**.

The SCANNER policy framework guides the development of network software that ensures that privacy and confidentiality are preserved in distributed research networks. The framework compares the SCANNER sites' privacy and security policy

requirements at the federal, state, and institutional level and can help to guide implementation of automated privacy and security services.

Other tools to facilitate governance of data and protect privacy include encryption, de-identification and anonymization mechanisms, and are integrated into many of the informatics strategies described in this brief. The forthcoming EDM Forum **Governance Toolkit** will include resources on the multiple dimensions of governance, including legal and regulatory concerns, data network architecture, governance bodies, IRBs, data properties, data sharing approaches and considerations, competitive marketplace situations, stakeholder engagement, and sustainability.

**Table 3. Population-Level Analytics**

	Title (Project)	Purpose and Function	Level of Modularity	Comments
Query & Explorers	*Shared Health Research Information Network (SHRINE) Custom Workbench (Pediatric Enhanced Registry)	Allows investigators to run cohort identification and feasibility queries on their own site's data or against the entire network	Plug & Play	Code is open source
	Research Data Explorer - RedX (WICER)	Supports querying of de-identified data in WICER Research Data Warehouse	Adaptable Approach	Approach is based upon New York-Presbyterian/Columbia University Medical Center data model
Distributed Analytic Tools	Grid Binary LOGistic REGression – GLORE (SCANNER)	Enables distributed analytics for multiple analysts at different institutions without exchanging personal health information (PHI)	Adaptable Approach	Open source
	Observational Cohort Event Analysis and Notification System – OCEANS (SCANNER)	A toolkit of statistical software designed to detect adverse event signals	Adaptable Approach	Open source

\* These tools are based on existing technologies or strategies that have been leveraged, expanded or improved by the projects (e.g., adopted or adapted for CER).

### Policy Implications and Next Steps

Arguably, the most efficient path forward to building learning health systems is to cultivate a community of interest and exchange ideas that encourage researchers to adopt and adapt existing tools rather than developing new resources from scratch. To achieve this goal of collaborative health science, there is a need for transparency and information sharing among key stakeholders interested in generating new tools and refining existing tools for research and QI. A key step is to lower the barriers to collaboration by developing open source communities that make it easy to develop, share, and test new tooling against reference datasets, or invite opportunities for collaborators to use and compare tools against proprietary datasets, even when data is not exchanged.

There is also a need to incentivize collaboration among those doing this important work. For example, encouraging grant making agencies to support (or require) individuals' contributions to open science, or encouraging promotion and tenure committees to value such work.. As part of these efforts, new strategies to fully credit developers' work is an important policy consideration moving forward.

With AHRQ's support, the EDM Forum is fully engaged in building a scientific collaborative that can advance the infrastructure for health research and QI, as well as improve the science and outcomes of these initiatives. As one example of these efforts, an initial discussion of collaboration will focus on researcher and developers' perspectives about contributing to open code, data, and tools. An upcoming, freely available webinar, "Cultivating Collaboration – Sharing Data, Code, and Tools to Accelerate the Science of Healthcare" will take place on August 29, 2013.. The webinar will feature Anthony D'Amico (Kaiser Family Foundation), Xiaoqian Jiang (University of California, San Diego), Daniella Meeker (RAND Corporation), and Fred Trotter (Cautious Patient Foundation and DocGraph) discussing new models of collaboration and ways the multi-disciplinary community interested in using ECD for research and QI can contribute.

On an ongoing basis, the EDM Forum community is invited to participate in further dialogue and exchange. Comments and contributions may be submitted to the EDM Forum eRepository at any time. Suggestions for open access tools in the domains of interest are welcome.

### Conclusion

Current investments to build new and improved infrastructure for ECD are focused on improving the accuracy, efficiency, and timeliness of conducting CER, PCOR, and QI, with the ultimate objective of improving patient care and outcomes. The 31 informatics platforms and tools discussed in this brief represent a set of innovative approaches to facilitate discovery and QI in support of researchers, clinicians, and patients. In light of the rapid evolution in healthIT, there is a need for collaborative venues to exchange promising technologies and support further innovation.

There has been great progress to date facilitating rapid, efficient, and meaningful evidence generation in a learning health system. Yet there is much to be done to achieve the goals of a functional, accessible learning system for all. While the PROSPECT, DRN, and Enhanced Registry projects hearken a paradigm shift towards the use of large heterogeneous datasets from multiple clinical settings, these efforts are early in their lifecycle, and further validation and applications are required. Ongoing testing of these systems and tools, as well as their findings will undoubtedly raise new questions—both possibilities and concerns. In the short term, however, their primary contribution is to demonstrate the feasibility of these prototype systems to resolve traditional issues of data access and study design. Moving forward, these tools will facilitate (and necessitate) ongoing collaboration among key communities of interest to achieve their full potential and utility. With the ongoing support of the community, there is great potential for informatics tools for research and QI to advance health science to improve patient care and outcomes.



**Table 4: Provider, Researcher, or Patient-Facing Decision Support**

Title (Project)		Purpose and Function	Level of Modularity	Comments
Clinical/QI Decision Support Systems	Care Stratification Algorithm (Pediatric Enhanced Registry)	Assigns a risk score to patients based on a selected set of clinical variables	Adaptable Approach	Approach is generalizable to all chronic conditions and the algorithm is freely available
	*Clinical Decision Support Software (Indiana PROSPECT)	Software that provides the capacity to alert clinicians to potential study candidates	Plug & Play	Open source version forthcoming
	Data Quality Reports (Pediatric Enhanced Registry)	Provides automated reports delivered to care centers via a centralized web portal, measuring the quality of data that are being entered into the registry	Hosted Platform	Reports generated for sites within the network
	*Population Management Tools (Pediatric Enhanced Registry)	Provides automated reports that allow staff to view patient population by grouping patients according to care stratification variables, clinical reporting measures, demographics, and medication use	Adaptable Approach	Design is freely available and the approach is adaptable to all chronic conditions
	*Pre-Visit Planning Tools (Pediatric Enhanced Registry)	Provides automated reports contain basic patient information, select data from past visits as well as care stratification scores, and considerations (recommendations) for medication dosing and lab ordering	Adaptable Approach	Design of the tool is freely available. Approach is adaptable to all chronic conditions
	PROs in Practice (CERTAIN)	An interactive data platform for real-time use that assists patient and clinician communication and decision-making at the point-of-care, and provides benchmarking data	Hosted Platform	Developed within the CERTAIN network
Research Decision Support Systems	Integrated Model for Patient Care and Clinical Trials (IMPACT) - Research Decision Support Software (WICER)	Software that integrates clinical research workflow in patient care settings through a set of research decision support functions and features	Plug & Play	Design of the tool is generalizable and portions of the code can be reused
	*Linkage Database and Consent Management Tool (Pediatric Enhanced Registry)	Web-based application that manages study participants and their protected health information	Adaptable Approach	Built on mapping for study-specific identifiers and patient identifiers for specific care centers
Patient Portals/ Navigators	Electronic Tailored Infographics for Community Engagement, Education, and Empowerment – EnTICE3 (WICER)	Approach supports creation of tailored visualizations based upon data collected from survey participants	Adaptable Approach	Approach and design of the tool are generalizable
	*Patient Activation Tools: Patient Status Tool (Pediatric Enhanced Registry)	Provides patients with a snapshot of their current disease activity and history	Adaptable Approach	Design of the tool is generalizable and portions of the code can be reused

\* These tools are based on existing technologies or strategies that have been leveraged, expanded or improved by the projects (e.g., adopted or adapted for CER).

### About AcademyHealth

AcademyHealth is the leading national organization serving the fields of health services and policy research and the professionals who produce and use this important work. Together with our members, we offer programs and services that support the development and use of rigorous, relevant and timely evidence to increase the quality, accessibility, and value of health care, to reduce disparities, and to improve health. A trusted broker of information, AcademyHealth brings stakeholders together to address the current and future needs of an evolving health system, inform health policy, and translate evidence into action.

### About the EDM Forum

The Electronic Data Methods (EDM) Forum is a three-year grant from the Agency for Healthcare Research and Quality (AHRQ) to advance the national dialogue on the use of electronic clinical data for the conduct of comparative effectiveness research (CER), patient-centered outcomes research (PCOR), and quality improvement (QI). The EDM Forum facilitates exchange and collaboration between eleven AHRQ-funded projects, including: the Prospective Outcome Systems using Patient-specific Electronic data to Compare Tests and therapies (PROSPECT) studies; the Scalable Distributed Research Networks for CER, the Enhanced Registries for QI and CER, as well as other relevant health information technology (healthIT) initiatives. The EDM Forum and the research projects connected to the EDM Forum are funded through the American Recovery and Reinvestment Act of 2009 (ARRA).

## Acknowledgements

AcademyHealth acknowledges the Agency for Healthcare Research and Quality (AHRQ) for its support of this work. The EDM Forum is supported by AHRQ through the American Recovery & Reinvestment Act of 2009, Grant U13 HS19564-01. AHRQ's mission is to improve the quality, safety, efficiency, and effectiveness of health care for all Americans. As 1 of 12 agencies within the Department of Health and Human Services, AHRQ supports research that helps people make more informed decisions and improves the quality of health care services. For more information, visit [www.ahrq.gov](http://www.ahrq.gov).

## Suggested Citation

AcademyHealth. "Informatics Tools and Approaches To Facilitate the Use of Electronic Data for CER, PCOR, and QI: Resources Developed by the PROSPECT, DRN, and Enhanced Registry Projects," EDM Forum, August 2013.

## References

1. "Building Electronic Clinical Data Infrastructure to Improve Patient Outcomes: CER Project Profiles," EDM Forum, AcademyHealth, April 2013.
2. US Department of Health and Human Services. Comparative Effectiveness Research Funding [HHS Recovery Act web site]. June 2010. Available at: <http://www.hhs.gov/recovery/programs/cer/index.html>. Accessed August 31, 2011.
3. Navathe AS, Clancy C, and Glied S. Advancing Research Data Infrastructure for Patient-Centered Outcomes Research. *JAMA*. 2011;306(11):1254-1255. doi:10.1001/jama.2011.1341.
4. IOM: <http://www.iom.edu/Reports/2011/Digital-Infrastructure-for-a-Learning-Health-System.aspx>
5. EDM Forum, op. cit.
6. Department of Health and Human Services, op. cit.
7. Clancy, op. cit.
8. IOM, op. cit.
9. Sittig DF, Hazlehurst BL. Informatics grand challenges in multi-institutional comparative effectiveness research. *JCER*. 2012 Sep; 1(5): 373-376.
10. Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research. *Med Care*. 2013 Jul; Suppl:S30-37.
11. Agency for Healthcare Research and Quality. ARRA-AHRQ Recovery Act 2009 Limited Competition: PROSPECT Studies: Building New Clinical Infrastructure for Comparative Effectiveness Research (R01) [PROSPECT Funding Opportunity Announcement web site]. December 8, 2009. Available at: <http://grants.nih.gov/grants/guide/rfa-files/RFA-HS-10-005.html>. Accessed November 1, 2011.
12. Agency for Healthcare Research and Quality. ARRA OS Recovery Act 2009 Limited Competition: Enhanced Registries for Quality Improvement and Comparative Effectiveness Research (R01) [Enhanced Registries Funding Opportunity Announcement web site]. January 21, 2010. Available at: <http://grants.nih.gov/grants/guide/rfa-files/RFA-HS-10-020.html>. Accessed November 1, 2011.
13. Agency for Healthcare Research and Quality. ARRA OS: Recovery Act 2009 Limited Competition: Scalable Distributed Research Networks for Comparative Effectiveness Research (R01) [Scalable Distributed Research Networks Funding Opportunity Announcement web site]. January 21, 2010. Available at: <http://grants.nih.gov/grants/guide/rfa-files/RFA-HS-10-015.html>. Accessed November 1, 2011.
14. PCMag.com. Definition of: plug and play [PCMag.com Encyclopedia web page]. Available at: <http://www.pcmag.com/encyclopedia/term/49389/plug-and-play>. Accessed July 1, 2013.
15. PCMag.com. Definition of: hosted environment [PCMag.com Encyclopedia web page]. Available at: <http://www.pcmag.com/encyclopedia/term/44374/hosted-environment>. Accessed July 1, 2013.
16. Sittig and Hazlehurst 2012, op. cit.
17. Ibid
18. Embi PJ, Kaufman SE, Payne PR. Biomedical informatics and outcomes research: enabling knowledge-driven health care. *Circulation* 2009;120:2393-9.
19. Linda Null; Julia Lobur (2006). The essentials of computer organization and architecture (2nd ed.). Jones & Bartlett Learning. pp. 33,179-181. ISBN 978-0-7637-3769-6.
20. R. Fagin, P. Kolaitis, R. Miller, and L. Popa. "Data exchange: semantics and query answering." *Theoretical Computer Science*, 336(1):89-124, 2005.
21. Sittig and Hazlehurst 2012, op. cit.
22. Kahn MG, Batson D, Schilling LM. Data model considerations for clinical effectiveness researchers. *Med Care*. 2012 Jul;50 Suppl:S60-67.
23. Observational Medical Outcomes Partnership [OMOP website]. 2011. Available at: <http://omop.fnih.org/>
24. Sittig DF, Hazlehurst BL, Brown J, et al. A survey of informatics platforms that enable distributed comparative effectiveness research using multi-institutional heterogeneous clinical data. *Med Care*. 2012 Jul;50 Suppl:S49-59.
25. Agency for Healthcare Research and Quality. ARRA OS: Recovery Act 2009 Limited Competition: Scalable Distributed Research Networks for Comparative Effectiveness Research (R01) [Scalable Distributed Research Networks Funding Opportunity Announcement web site]. January 21, 2010. Available at: <http://grants.nih.gov/grants/guide/rfa-files/RFA-HS-10-015.html>. Accessed November 1, 2011.
26. About Open mHealth [Internet]. Open mHealth. 2012 [cited 2013 Feb 25]. Available from: <http://openmhealth.org/about/>
27. Berner, Eta S., ed. *Clinical Decision Support Systems*. New York, NY: Springer, 2007. Available at: <http://www.openclinical.org/dss.html>
28. Torgan, Carol (November 6, 2009). "The mHealth Summit: Local & Global Converge". [caroltorgan.com](http://caroltorgan.com)
29. Nass SJ, Levitt L, and Gostin LO, ed. 2009. *Beyond the HIPAA Privacy Rule*. Washington, DC: Institute of Medicine.
30. IOM (Institute of Medicine). 2011. *Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary*. Washington, DC: The National Academies Press.

## Appendix A: Descriptions Of Informatics Platforms And Tools

### Table 1: Person-Level Data Collection

#### \*CareWeb Framework Clinical Plugins (Indiana PROSPECT)

The CareWeb framework is a reusable application framework that is intended for open source use. Clinical (e.g., results retrieval or order entry) can be inserted. CareWeb is built on a development platform called ZK, an open source web development framework that leverages AJAX and traditional Java. ZK allows implementation of standard web UI elements such as HTML/JavaScript/CSS and jQuery, but also has custom “ZUML” components that are built right into the framework.

#### \*Enhanced eMR-ABC software (Indiana PROSPECT)

A user-friendly, web-based electronic medical record system designed to work hand in hand with the Aging Brain Care model. The system enhances the existing infrastructure to capture richer clinical data and enhance the ability to incorporate patient outcomes, improve capture of device utilization data, incorporate genomic and other high throughput results, and identify patient for recruitment and enrollment technologies. For more information, visit <http://www.agingbraincare.org/tools/emr-abc/> and see <http://repository.academyhealth.org/egems/vol1/iss1/8>.

#### \*Open Source EHR-Based Registry (Pediatric Enhanced Registry)

The EHR module allows for centers to enter data-in-once from clinical encounters via a web form and have data stored in center-specific database. A separate module allows for centers to upload registry data that has been collected elsewhere. For more information, visit <https://enhancedregistry.org/>.

- *i2b2 Forms Cell* – allows for web-based data capture of clinical information and subsequent transfer to a center-specific i2b2 database. The form definitions are based on a customized version of the CDISC Operation Data Model (<http://www.cdisc.org/odm>), which was developed to allow for the representation and exchange of clinical research data and metadata. The i2b2 Forms Cell is open source. More information can be found here: <https://community.i2b2.org/wiki/display/CCHMC/Home>
- *Procedures to extract EHR data – database scripts (SQL stored procedures)* that allow data to be extracted from the EHR and put into a format suitable for upload into the registry (through the EHR upload module). The approach is adaptable, but due to Intellectual Property rules, the scripts can only be shared among customers of that vendor.
- *EHR upload module* – a secure web portal that allows users to upload a flat file (csv format) containing data that have been extracted from the EHR. Once a file has been uploaded, it is processed, which includes checking the form for errors (e.g.,

whether a patient exists in the registry, whether each visit has a date) and replacing the EHR-based identifiers with registry identifiers. After this swap has occurred, the data are transferred to the registry using i2b2 Forms Cell web services. The approach is adaptable to other registry projects.

#### Tablet App-Based and Web-based Data Collection (WICER)

A community-health survey adapted for primary data collection using tablets (iPads). The project also developed novel security and encryption strategies for cloud-based data storage approaches. For more information, visit <http://wicer.org/> and see <http://repository.academyhealth.org/egems/vol1/iss1/7>.

#### \*Patient Activation Tools: Inter-Visit Planning App (Pediatric Enhanced Registry)

The Inter-Visit Planner App is a package of electronic tools that use patient-specific data to enhance patient activation and encourage productive patient-clinician interactions. The mobile app designed to assess the patient’s symptoms and well-being in between clinic visits as well as provide a mechanism for parents to document any non-urgent questions that they wish to ask their care team at the next clinic visit. The code is not open source, but the approach is applicable to other chronic conditions and the team is interested in exploring opportunities with other partners. For more information, visit <http://c3nproject.org/innovations/patient-engagement-community/e3-healthcare-study>

#### Patient Voices Project Survey Center (CERTAIN)

The Center supports outbound survey distribution for point-of-care and long-term direct-to-patient follow-up for healthcare quality improvement initiatives, research studies, and non-hypothesis driven technology surveillance. The Center uses four participant contact methods (postal mail, email, telephone or Short Message Service [SMS] text message) and four survey response formats (internet-based/web survey, paper-and-pencil, live operator phone interview, automated phone interview). Eight combinations of contact methods and survey response format are in use: 1) Postal mail to paper-and-pencil; 2) Postal mail to web; 3) Postal mail to automated phone interview; 4) email to web; 5) email to automated phone interview; 6) telephone to live operator phone interview; 7) telephone to automated phone interview; 8) SMS text message to web. Surveys are customized with relevant project logos and information to increase participant recognition and response rate. The Center can distribute surveys on any protocol (e.g. at any time interval, for any length of time, in designated formats) and the report back is also flexible, options include: direct data download and transfer, access to a reporting portal for investigator driven ad-hoc summary reports, or full support in survey results analysis and reporting. For more information, visit <http://depts.washington.edu/chaseall/index.php>.

\* These tools are based on existing technologies or strategies that have been leveraged, expanded or improved by the projects (e.g., adopted or adapted for CER).

**\*Caradigm Amalga © (CERTAIN)**

The platform incorporates semi-automation of EMR abstraction and allows flexible data aggregation capabilities to support patient-centric analytics and a centralized repository with near real-time data access. The platform leverages open source NLP algorithms developed by the project validating automated EMR extraction to registry. For more information, visit <http://www.caradigm.com/en-us/products/caradigm-intelligence-platform/>.

**CER Hub Collaborative web-based informatics platform (CER Hub)**

This platform will provide a means to propose, develop, and conduct multi-institutional (or single site) studies using EMR data. Components that enable this are:

- A set of policies that identify how to create and conduct a new CER Hub study or how to participate as CER Hub data provider -- available now at [www.cerhub.org](http://www.cerhub.org)
- A common data model (the Clinical Research Document [CRD] as implementation of HL7 Clinical Document Architecture ([CDA] 2.0) -- available October 2013
- A general purpose tool for manipulating CRD documents (the emrAdapter) -- available to any CER Hub study
- A web-based resource for building and evaluating MediClass applications that apply natural language processing to classify CRD documents -- available to any CER Hub study
- MediClass applications for download from the CER Hub application library, or developed for new topics by study team – available to any CER Hub study

**Table 2: Data Access, Exchange, And Aggregation Data Cleaning Tool (Pediatric Enhanced Registry)**

The data cleaning tool converts the raw data from the registry into a research-grade database along with its codebook. This “reproducible research” tool is written in R, and allows for codebook and database updates by running the tool on as frequent a basis as desired. The approach is adaptable to other projects.

**\*Diabetes DataLink (SUPREME-DM)**

A multisystem distributed database that captures and standardizes detailed demographic and longitudinal electronic clinical data for conducting CER and surveillance in the treatment and prevention of diabetes mellitus. The network utilizes a **data repository with structured data (based on the HMORN VDW) to conduct analyses**. Intended users could be: IRBs, Principal Investigators, and Project Managers when creating/requesting an umbrella IRB approval for multi-site studies that use a (in our case distributed) database resource for research. New stud-

ies (sub-studies) are submitted as modifications to the umbrella approval which keeps the required protections and oversight in place while accelerating the process to obtain approvals for this research. For more information, visit <http://www.supreme-dm.org/Cores.html>.

**Research Data Warehouse (WICER)**

The WICER research data warehouse includes de-identified data including clinical data from the New York-Presbyterian/ Columbia University Medical Center clinical data warehouse linked to WICER survey for participants who agreed to data linkage. Survey participants comprise only about 3% of the patients in the clinical data warehouse. The linkage and de-identification approaches can be adapted by others. For more information visit <http://www.wicer.org>

**Reusable OMOP and SAFTINet Interface Adaptor - ROSITA (SAFTINet)**

ROSITA is a fully packaged virtual machine that performs five critical tasks for preparing EHR data for multi-institutional data sharing: (1) syntactic and semantic data harmonization into the OMOP common data model and terminology, (2) management of personal identifiers to expose only HIPAA-compliant limited data sets, (3) clear-text and encrypted record linkage, (4) data quality assessment and reporting, and (5) site-level control of data elements that are made available to the research network. ROSITA has been successfully deployed at multiple locations with different clinical data systems. For more information, visit <http://saftinet.net/rosita> and see <http://repository.academyhealth.org/egems/vol1/iss1/11>.

**Sleep Medicine SharePoint Platform (COMET)**

The Sleep Medicine platform supports numerous applications such as comparative effectiveness trials, patient portals, registries and others. The platform is composed of two major subsystems: 1) an open source data capture and processing pipeline which includes instrument libraries, study management tools, and business intelligence analysis features and 2) a presentation and collaboration layer based on Microsoft SharePoint which provides content management, social and collaborative features, and rich data visualization tools. For more information, visit <http://comet.stanford.edu>.

**\*PopMedNet Datamarts: Obesity and ADHD Modules (SPAN) and Asthma Module (PEAL)**

PopMedNet is a distributed research network architecture that permits menu-driven querying and the execution of distributed programs that access data in a common data model. Hypothesis-generating tools are incorporated in the interface to gauge the feasibility of conducting future CER. For more information, visit [www.popmednet.org](http://www.popmednet.org). Two of the projects have developed condition-specific datamarts onto the platform:



- **Obesity and ADHD Data Marts** The development of standardized data marts for obesity and ADHD that are based on standard tables and fields in the HMO Research Network's Virtual Data Warehouse (VDW). It is the concept of the VDW, the thought given to the tables and fields that it should contain, and the extract, transform, and load (ETL) procedures to convert data from its sources to an accessible research database that has made the creation of the data marts and the application of the query tool to these data marts possible.
- **PEAL Data Warehouse** A set of federated datasets that is structured similarly to the HMORN VDW and includes standard VDW files on a cohort with asthma or other lung diseases from the 6 participating sites. Data on the asthma cohort are held in a central repository and other data are held locally at each site. This data warehouse can be used for preliminary analyses preparatory to research, as well as new research projects, with appropriate IRB approvals. The data dictionary and ICD-9 code sets to identify asthma and other lung diseases are also available as resources. The data dictionary and codes are considered open source.

#### **\*HMORN Virtual Data Warehouse Common Data Model (SUPREME-DM, SPAN, CER Hub, PEAL Network)**

The SUPREME-DM, SPAN, CER Hub, and PEAL Network leverage the HMO Research Network Virtual Data Warehouse (VDW) common data model. The HMORN VDW is a series of dataset standards and automated processes in place at implementing sites that allow SAS programs written at one HMORN Site to be run against other VDW sites quickly and with a minimum of site-specific customization. It is 'virtual' in the sense that there is no centrally located store of data against which data from all sites can be touched in a single run. For more information, visit [http://www.hmoresearchnetwork.org/resources/toolkit/HMORN\\_CollaborationToolkit.pdf#4](http://www.hmoresearchnetwork.org/resources/toolkit/HMORN_CollaborationToolkit.pdf#4).

#### **OMOP Common Data Model Version 4 (SAFTINet and SCANNER)**

The Common Data Model (CDM) standardizes the format, and the Standard Vocabulary contains the content or coding schemes used across disparate data sources and types. Version 4.0 supports health economics use cases and medical treatment outcome studies, including medical device safety, comparative effectiveness, and healthcare quality. Created in collaboration with the Observational Medical Outcomes Partnership and the SAFTINet and SCANNER projects. For more information, visit <http://omop.fnih.org/CDMvocabV4> and <http://scanner.ucsd.edu/about/semantic-interoperability>.

#### **\*Translational Informatics and Data Management Grid - TRIAD (SAFTINet)**

The SAFTINet network deploys TRIAD grid technology, the OMOP common data model, detailed technical documentation,

and custom software for data harmonization to facilitate data sharing in collaboration with stakeholders in the care of safety net populations. TRIAD, an domain-agnostic derivative of the caGrid data sharing platform developed under the auspices of the NCATS-funded CTSA program, is a stateful and secure service oriented architecture that is used as the trusted communication and grid networking fabric for SAFTINet.

The Biomedical Informatics Program at The Ohio State University (OSU) hosts the core TRIAD services and functionality on which SAFTINet relies. It is also possible for a network to implement and host TRIAD independently. TRIAD's middleware system is designed to create a loosely coupled, yet highly interoperable grid service oriented architecture (SOA). It has been adopted as the basis for the TRIAD Community grid system. The TRIAD System provides two primary classes of services: (1) Security and Indexing Services and (2) Data Services.

#### **Table 3: Population-Level Analytics**

##### **\*Shared Health Research Information NETWORK (SHRINE) Custom Workbench (Pediatric Enhanced Registry)**

A "virtual" SHRINE network was implemented among the i2b2 databases hosted at the data coordinating center in the Pediatric Enhanced Registry. The custom workbench allows investigators to run cohort identification and feasibility queries on their own Care Center's data (patient-level, using i2b2) or against the entire Network (center-level, using SHRINE). The code for both SHRINE and i2b2 are open source (<https://community.i2b2.org/wiki/dashboard.action>). The Cincinnati Children's Hospital Medical Center custom workbench can be found here: <https://bmi.cchmc.org/svn/i2b2/i2b2/public/>

##### **Research Data eXplorer – RedX (WICER).**

A front-end viewer to de-identified data in the Research Data Warehouse. RedX allows researchers to browse patient data to look for specific data elements that can be used to generate research queries (identified as a critical component of query creation through years of experience providing mediated queries). The current version is tied to New York Presbyterian Hospital data model, but could be adapted to other data models. For more information, visit <http://www.wicer.org>.

##### **Grid Binary LOGistic REGression – GLORE (SCANNER)**

An open source distributed analytics tool that enables researchers to build an accurate logistic regression model in a privacy-preserving manner without the data leaving individual sites. All sites send local intermediary analysis results to the server, the server integrates these results, then the server sends global results back to the sites, where the sites re-compute local intermediary analysis results. This cycle repeats until the estimated parameters converge, which usually takes less than 15 iterations. The final model is as accurate as if it had been constructed from a centralized data repository. For more information, visit <http://dbmi-engine.ucsd.edu/webglore2/> and see:



Wu Y, Jiang X, Kim J, Ohno-Machado L. Grid Binary Logistic Regression (GLORE): building shared models without sharing data. *J Am Med Inform Assoc.* 2012 Sep-Oct;19(5):758-64.

#### **Observational Cohort Event Analysis and Notification System – OCEANS (SCANNER)**

An open source statistical analysis toolkit that contains different statistical methods that can be used to detect adverse event signals. Selection of the statistical method would depend on factors such as how a researcher wants to handle repeated measurement error or alerting thresholds. The current version of OCEANS includes methods for proportional difference analysis and sequential probability ratio tests, e.g., propensity score matched proportional difference and risk adjusted sequential probability ratio test. For more information, visit <http://sourceforge.net/projects/oceans/>.

#### **Table 4: Provider, Researcher Or Patient-Facing Decision Support**

##### **Care Stratification Algorithm (Pediatric Enhanced Registry)**

The Pediatric Enhanced Registry care stratification algorithm was developed to assign a risk score to patients based on a selected set of clinical variables. Includes measures like steroid use, disease status, psychosocial issues, growth and nutrition status. The approach is generalizable to all chronic conditions and the algorithm is freely available.

##### **\*Clinical Decision Support Software (Indiana PROSPECT)**

This CDS software, that is intended for open source use, provides the capacity to alert clinicians to potential study candidates and allow them to confirm that a patient is eligible and obtain consent for the patient to be approached by a research assistant.

##### **Data Quality Reports (Pediatric Enhanced Registry)**

The Pediatric Enhanced Registry data quality reports are automated and delivered to centers via a centralized web portal. The data quality reports measure the quality of the data that are being entered into the registry for the patients and patient visits. In addition, the reports provide information on areas within the Care Center's registry data where the data have not been entered completely (e.g., an exception report).

##### **\*Population Management Tools (Pediatric Enhanced Registry)**

Pediatric Enhanced Registry population management tools provide automated reports that allow Care Center staff to view their patient population by grouping patients according to care stratification variables, clinical reporting measures, demographics, and medication use. For each summary measure, a Care Center can see both the number and the percentage of patients within each grouping for that specific measure. The Care Center can then drill-down further on a particular grouping and receive the listing of patients that match the criteria. The listing of patients, along with high-level summary information, can be printed, or the center can select a specific patient and the Pre-

Visit Planning (PVP) report will be generated. For the patient listing the Care Center, the patient names and MRN will display for those patients that have been populated. The design is freely available and the approach is adaptable to all chronic conditions. For more information: <http://c3nproject.org/innovations/qi-and-chronic-illness-care-support/population-management-automated-reporting>

##### **\*Pre-Visit Planning Tools (Pediatric Enhanced Registry)**

The Pediatric Enhanced Registry pre-visit planning tools generate automated reports contain basic patient information, selected data from the past 8 visits, as well as care stratification scores, and considerations (recommendations) for medication dosing and lab ordering. Other considerations are also included based on the patient's registry values to help aid the center in preparing for a patient's visit. The design is freely available and the approach is adaptable to all chronic conditions. For more information: <http://c3nproject.org/innovations/qi-and-chronic-illness-care-support/pre-visit-planning>

##### **PROs in Practice (CERTAIN)**

The CERTAIN PRO in practice platform is an interactive data platform that facilitates real-time use of PROs to assist patient and clinician communication and in decision-making at the point-of-care, in addition to developing benchmarking data for public access and use. For more information, visit [http://www.becertain.org/hospitals/patient\\_voices/pros](http://www.becertain.org/hospitals/patient_voices/pros).

##### **Integrated Model for Patient Care and Clinical Trials (IMPACT) – Research Decision Support (WICER)**

The Integrated Model for Patient Care and Clinical Trials (IMPACT) provides research decision support in the clinical research setting. Using a multi-user constraint satisfaction and resource optimization algorithm, IMPACT automatically synthesizes temporal availability of various research resources and recommends the optimal dates and times for pending research visits. This tool helps to schedule a patient's visit in coordination with research activities, improving efficiency by getting data for research purposes during a clinical visit. For more information visit <http://www.wicer.org> and see: Weng C, Li Y, Berhe S, Boland MR, Gao J, Hruby GW, Steinman RC, Lopez-Jimenez C, Busacca L, Hripscak G, Bakken S, Bigger JT. An Integrated Model for Patient Care and Clinical Trials (IMPACT) to support clinical research visit scheduling workflow for future learning health systems. *Journal of Biomedical Informatics.* 2013 May 16 – eprint ahead of print.

##### **\*Linkage Database and Consent Management Tool (Pediatric Enhanced Registry)**

The Pediatric Enhanced Registry includes a web-based application that was developed to manage study participants and their protected health information (PHI) by Care Center. The application is accessed through the main registry, however, the PHI is stored in an encrypted consent database separate from the registry data.

Participants can be added and maintained within the consent management system either individually or through a bulk upload process. The application stores demographic information on participants. For the pediatric enhanced registry, that information is MRN, name, date of birth, zip code (optional), e-mail address (optional), registry ID and center ID. The application tracks a patient's consent status (used to determine who can be included in a research dataset) and can fire alerts to the study staff when the patient's consent expires (when they turn 18, for instance). Once a patient's record has been created in the application, patients can be registered in either a paper or e-consent workflow (Care Centers can choose on a patient-by-patient basis). With e-consent, a parent (or patient, depending on age) is taken to a web portal where they are presented with a Care Center-and-age-specific consent form to which they can consent/assent to participate in the study.

Because the consent management system contains a mapping between study-specific identifiers and patient identifiers, another major use case of the tool is to act as a service that can be queried to replace registry IDs with patient names or MRNs (or vice versa) to produce clinical pre-visit planning and population management reports. This replacement of ID with Name / MR Number occurs in real-time and is only available for viewing when reports are run by the users at specific Care Centers.

### **Electronic Tailored Infographics for Community Engagement, Education, and Empowerment (WICER)**

Electronic Tailored Infographics for Community Engagement, Education, and Empowerment (EnTICE<sup>3</sup>) is designed to support creation of tailored visualizations based upon data collected from survey participants. The resulting infographics may be at the individual level or at the community level. The approach was motivated by the desire to return data to survey participants with relatively low levels of health literacy. The initial infographics are static and designed for display on a community portal, but the intent is to move toward the creation of interactive visualizations. For more information visit <http://www.wicer.org> and see: Arcia A, Bales ME, Brown III W, Co Jr MCC, Gilmore M, Lee YJ, Park CS, Prey J, Velez M, Woollen J, Yoon S, Kukafka R, Merrill JA, Bakken S. Method for the development of data visualizations for community members with varying levels of health literacy. Proceedings of Fall AMIA Symposium (in press).

### **\*Patient Activation Tools: Patient Status Tool (Pediatric Enhanced Registry)**

The patient status tool (PST) is used to provide patients with a snapshot of their current disease activity as well as their recent history. It consists of a user-friendly version of the patient's status (labs, height/weight, PGA, etc.) along with a set of system-generated talking points for the patient's next visit. The PST is delivered to parents 5-7 days before a visit, at which point the patient's clinical data are pulled from the registry and the PST is delivered to the parent and available for review online through a secure website. The design of the tool is generalizable and portions of the code can be reused. For more information, visit <http://c3nproject.org/innovations/patient-engagement-community/e3-healthcare-study>.