METHOD ARTICLE

Development of an informatics system for accelerating biomedical research. [version 1; peer review: 1 approved with reservations, 1 not approved]

Vivek Navale¹ , Michele Ji¹, Olga Vovk², Leonie Misquitta³, Tsega Gebremichael³, Alison Garcia³, Yang Fann⁴, Matthew McAuliffe¹

¹Office of Intramural Research, Center for Information Technology, National Institutes of Health, USA, Bethesda, Maryland, 20892, USA

²General Dynamics Information Technology, Inc., Fairfax, Virginia, 22030, USA

³Sapient Government Services, Arlington, Virginia, 22201, USA

⁴Intramural IT and Bioinformatics Program, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, 20892, USA

V1 First published: 14 Aug 2019, 8:1430
<https://doi.org/10.12688/f1000research.19161.1>
 Latest published: 13 Jul 2020, 8:1430
<https://doi.org/10.12688/f1000research.19161.2>

Abstract

Biomedical translational research can benefit from informatics system that support the confidentiality, integrity and accessibility of data.

Such systems require functional capabilities for researchers to securely submit data to designated biomedical repositories. Reusability of data is enhanced by the availability functional capabilities that ensure confidentiality, integrity and access of data. A biomedical research system was developed by combining common data element methodology with a service-oriented architecture to support multiple disease focused research programs. Seven service modules are integrated together to provide a collaborative and extensible web-based environment. The modules - Data Dictionary, Account Management, Query Tool, Protocol and Form Research Management System, Meta Study, Repository Manager and globally unique identifier (GUID) facilitate the management of research protocols, submitting and curating data (clinical, imaging, and derived genomics) within the associated data repositories. No personally identifiable information is stored within the repositories. Data is made findable by use of digital object identifiers that are associated with the research studies. Reuse of data is possible by searching through volumes of aggregated research data across multiple studies. The application of common data element(s) methodology for development of content-based repositories leads to increase in data interoperability that can further hypothesis-based biomedical research.

Open Peer Review

Approval Status

	1	2
version 2		
(revision)		
13 Jul 2020		
version 1		
14 Aug 2019		

1. **Hyeoneui Kim** , Duke University, Durham, USA

2. **Timothy W. Clark** , University of Virginia, Charlottesville, USA

Any reports and responses or comments on the article can be found at the end of the article.

Keywords

Informatics system, Biomedical repository, Translational Research, FAIR

This article is included in the **Data: Use and Reuse** collection.

Corresponding author: Vivek Navale (Vivek.Navale@nih.gov)

Author roles: **Navale V:** Conceptualization, Formal Analysis, Investigation, Methodology, Project Administration, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Ji M:** Methodology, Software; **Vovk O:** Data Curation, Investigation; **Misquitta L:** Data Curation, Investigation; **Gebremichael T:** Software, Visualization; **Garcia A:** Writing – Review & Editing; **Fann Y:** Project Administration, Writing – Review & Editing; **McAuliffe M:** Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2019 Navale V *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The author(s) is/are employees of the US Government and therefore domestic copyright protection in USA does not apply to this work. The work may be protected under the copyright laws of other jurisdictions when used in those jurisdictions.

How to cite this article: Navale V, Ji M, Vovk O *et al.* **Development of an informatics system for accelerating biomedical research. [version 1; peer review: 1 approved with reservations, 1 not approved]** F1000Research 2019, 8:1430 <https://doi.org/10.12688/f1000research.19161.1>

First published: 14 Aug 2019, 8:1430 <https://doi.org/10.12688/f1000research.19161.1>

Introduction

Translational Medicine (TM) seeks to develop new treatments for diseases with insights towards the improvement of global health. To achieve this vision, understanding what and why interventions work, and how they can be scaled to benefit the entire population, depends on successful translational biomedical research and data lifecycle management. The process of TM is time consuming with translational barriers, from basic research to clinical application, bedside to community use, and from community to policy-making decisions. In overcoming the translational barriers, that biomedical informatics platforms reduce the time it takes for basic research to result in clinical applications¹. To date, biomedical platforms for translational research have been developed for one or more purposes - (a) management of multi-dimensional heterogeneous data, (b) dissemination of knowledge generated during translational research, (c) testing of analytic approaches for data pipelines, and (d) application of knowledge-based systems and intelligent agents to enable high-throughput hypothesis generation².

Several biomedical informatics applications have been discussed in the literature; for example, Research Electronic Data Capture (REDCap) is a software tool for collecting, storing, creating project specific databases for dissemination of clinical and translational research data^{3,4}. The informatics for integrating Biology and the Bedside (i2b2) system allows researchers to find cohorts of patients that fit specific profiles⁵. Access to chemical, 'omics' and clinical data, with capabilities to investigate genetic and phenotypic relationships for cohorts of patients is supported by transSMART platform^{6,7}. Analyses of large complex datasets with bioinformatics and image analysis tools, cloud services, application programming interfaces (APIs), and data storage capabilities is supported by the CyVerse infrastructure^{8,9}. Software tools to collect, manage, and share neuroimaging data of different modalities, including magnetic resonance imaging (MRI), magnetoencephalography (MEG), and electroencephalogram (EEG) is available through the Collaborative Informatics and Neuroimaging Suite (COINS)^{10,11}. Also, large scale analysis of biological data can be carried out by web-based platforms¹². Within the National Institutes of Health (NIH), the Biomedical Translational Research Information System (BTRIS) has supported researchers to bring together data from the NIH Clinical Center and other institutes and centers¹³.

However, many disease focused research programs have faced data discoverability and integration challenges. For example, traumatic brain injury (TBI) research data was initially collected in different ways and by disparate systems making sharing and reusing of data problematic. Because of the wide variability in systems and databases, many types of TBI injuries were classified as the same class of injury, impeding development of targeted therapies for the disease. To overcome these barriers, the TBI community recommended use of the common data elements (CDE) methodology for the development of the Federal Interagency Traumatic Brain Injury Research (FITBIR)¹⁴.

A CDE is defined as a fixed representation of a variable collected within a specified clinical domain, that needs to be

interpretable unambiguously in human and machine-computable terms¹⁵. It consists of a precisely defined question with a specified format, with a set of permissible values as responses. Typically, CDE development for biomedical disease programs involves multiple steps - identification of a need for a CDE or group of CDEs, stakeholders and expert groups for CDEs selection, iterations and updates to initial development with ongoing input from broader community, with final endorsement of the CDEs by the stakeholder community for its usage and widespread adoption¹⁶.

Examples of CDEs use in various programs of clinical research include neuroscience¹⁷, rare diseases research¹⁸, and management of chronic conditions¹⁹. For clinical data lifecycle management, the use of CDEs provides a structured data collection process, which enhances the likelihood for data to be pooled and combined for meta-analyses, modeling, and post-hoc construction of synthetic cohorts for exploratory analyses²⁰. Investigators working to develop protocols for data collection can also consult the NIH Common Data Element Resource Portal for using established CDEs for disease programs²¹. Also, the feasibility of using common data elements and a data dictionary for the development of the National Database for Autism Research was shown earlier²².

Three key components comprise data dictionaries: data elements, form structures and eForms. A data element has a name, precise definition, and clear permissible values, if applicable. A data element directly relates to a question on a paper, electronic form (eForm) and/or field(s) in a database record. Form structures (FS) serve as the containers for data elements, and the eForms are developed by using FS. The data dictionary provides defined CDEs, as well as unique data elements (UDEs) for specific implementation of the BRICS instance. Reuse of CDEs is significantly encouraged, and in the case of FITBIR's data dictionary, it incorporates and extends the CDE definitions developed by the National Institute of Neurological Disorders and Stroke (NINDS) CDE Project¹⁵.

In this paper we demonstrate the application of CDE concept for developing a Biomedical Research Informatics Computing System (BRICS), providing functionalities that facilitate electronic submission of research data, validation, curation, and archival storage within program specific data repositories. Use of CDEs enhances data quality and consistency within the repositories that are important for advancing clinical and translational research.

Method

A high level overview of the informatics system architecture is provided in [Figure 1](#). The architecture is defined by the three layers - (a) Presentation Layer, (b) Application Layer, and (c) Data Layer. The Presentation Layer serves as the secure entry point to the BRICS portal. Various open source technologies and libraries, including Java Server Pages (JSP), jQuery, JavaScript libraries (e.g. such as Backbone.js, Asynchronous JavaScript), and XML are used to make web-pages interactive. This layer also includes Web Start applications: Global Unique

Figure 1. A schematic representation of the informatics system architecture.

Identifier (GUID) client, Validation and Upload tools, and Download and Image Submission tools, all of which run on users' machines.

The Image Submission Package Creation Tool leverages the 35 plus medical image file readers in the **Medical Image Processing Analysis and Visualization** (MIPAV) software (v 8.0.2), to make data interoperable, mapping of image header data onto the data elements in imaging form structures for submission to the Data Repository. MIPAV is an open-source software that can be used for image analysis, it is accessible on any Java-compatible platform, including Windows, Mac OS X, and Linux. Over 30 file formats commonly used in medical imaging, including DICOM and NifTI, and more than ten 3D surface mesh formats are supported by the software²³. It also supports multi-scale and multi-dimensional image research from various modalities including microscopy, computerized tomography (CT), positron emission tomography (PET), and MRI. Inclusion of the MIPAV tool with the BRICS provides capabilities for uploading image packages and image analysis, that is not conveniently available on other informatics systems²⁴.

The Application Layer is responsible for the logic that determines the capabilities of the BRICS modules and tools. Seven service modules within the Application Layer are integrated together to provide a collaborative and extensible web-based environment. These modules are the Data Dictionary (DD), Account Management, Query Tool, Protocol and Form

Research Management System (ProFoRMS), Meta Study, Repository Manager and GUID. To communicate and exchange information between the modules, **representational state transfer** (RESTful) Web services are used.

Additional information on the various service modules is available from the **BRICS site**.

The Data Layer consists of open source databases such as **PostgreSQL**, **Virtuoso** databases, file servers, and data persistence frameworks. The Virtuoso database is used to store the data accessed by the Query Tool and to store CDE metadata in the Data Dictionary data. The Repository module uses the PostgreSQL database to store and retrieve data. Also utilized are open-source libraries such as **Hibernate** and **Apache Jena** for storing and retrieve data from databases. The data layer is supported by the physical infrastructure located within the National Institutes of Health, and is certified at the Federal Information Security Modernization Act (FISMA) Moderate level²⁵, conforming to additional USA federal information standards^{26,27}.

To de-identify data, researcher's use the GUID tool (shown as a client in **Figure 1**) to assign a unique identifier for each study participant. The GUID is a random alphanumeric unique subject identifier (not generated from personally identifiable information (PII)). The PII fields that can be used as part of the hashing process include complete legal given (first) name of subject at birth, middle name (if available), complete legal family (last)

name of subject at birth, day of birth, month of birth, year of birth, name of city/municipality in which subject was born, and country of birth. The PII data is not sent to the GUID server but rather one-way encrypted hash codes are created and sent from the GUID client to the server (represented as a service module, [Figure 1](#)), allowing the PII to reside only on the researcher's site. A random number for each of the research participant is generated by the server and is returned to the researcher. In addition, the GUID server can be configured to support multi-center clinical trials and investigations that enroll research participants across various programs.

Information package preparation

Researchers are responsible for most of the data submission activities, which includes study FS approval, eForms review, curation, mapping of data elements, and providing associated study documentation.

Two routes of data submission are available for researchers to make data findable. One approach is by using the ProFoRMS tool ([Figure 2](#), stage 1) for clinical research work, scheduling subject visits, collecting data, adding new data, modifying previously collected data entries, and correcting discrepancies that are tracked and maintained in audit logs. The other mode is by using a generic data collection system (e.g. RedCap), validating with the BRICS data dictionaries and uploading the extracted data into the repository module ([Figure 2](#), stage 2). Both routes of data submission validate the submitted data using specific range and values from the data dictionaries for a BRICS instance.

The Validation Tool supports the data repository and ProFoRMS modules, by using CDEs with defined range and value metrics for data quality checks, to make data reusable. Once the data has

been validated and uploaded via the submission upload tool, data is stored in its raw form within the repository module in a database that can be accessed by the Query Tool ([Figure 2](#), stage 3).

User support is provided for data stewardship activities that include training and assistance to authorized users, for CDE implementation, data validation and submission to the repositories. Access is controlled by a Data Access Committee (DAC) that reviews studies for relevance to a specified BRICS instance (defined by the biomedical program). In addition, access to the system is role based and specific permissions are associated with roles such as PI, data manager, and data submitter.

During packaging of data, GUIDs are assigned to research subjects (patients), using the GUID client with the users responsible for storing PII data locally within their institutional systems. Data curation is carried out by identifying the available standard forms and CDEs in the Data Dictionary. In the event no corresponding CDEs are available, then the user can define the data elements and obtain approval during the submission process.

Information package storage and management

The data Repository module serves as a central hub, providing functionality for defining and managing study information and storing the research data associated with each study ([Figure 2](#), stage 3). Authorized investigators can submit data to a BRICS instance, organize one or many datasets into a single entity called a Study. In general terms, a 'Study' is a container for the data to be submitted, allowing an investigator to describe, in detail, any data collected, and the methods used to collect the data, which makes data accessible. By using the repository user interface, researchers can generate digital object identifiers (DOIs) for a study, which can be referenced in research articles.

Figure 2. Schematic representation of 1. Submission Information Package (SIP), 2 - Archival Information Package (AIP) preparation, 3 - storage of AIPs, and 4 - Dissemination Information Packages (DIP) access.

The repository module provides download statistics for specific studies, enabling the investigator to obtain information on their respective data that has been downloaded for other research activities, and overall increase data sharing and collaboration for additional research goals. Depending on the research studies, BRICS based repositories can host high throughput gene expression, RNA-Seq, SNPs, and sequence variation data sets (Figure 2, stage 3).

Data sharing

By default, the system assigns the sharing preference as 'private' where only users to that specific study can access the data. When the data is in the private state, the PI has the option to share data with specific collaborators (preferential sharing). After a certain period (defined by the data sharing policy for each BRICS instance), the data enters a new 'shared' state, which is accessible to the approved users.

Raw data is available for querying within 24 hours of data submission. For the data to be available via the Query Tool module, the raw data is processed through the 'NextGen Connect' tool (integrated interface engine) and Resource Description Framework (RDF) data interchange tool (Figure 2, stage 4). Shared data is available to all system users (approved by DAC) to search, filter, and download via the Query Tool functionality. The Query Tool offers three types of functionalities - (a) querying and filtering data, (b) data package downloads based on query, and (c) data package to the Meta Study module.

The Meta Study module is used for meta-analysis of the data as well as a collaboration tool between scientific groups. A Meta

Study contains findings from studies that can be aggregated by researchers to conduct additional analysis. The Query tool can also support the statistical computing language R as well as structured visualization of data (Figure 2, stage 4).

Result

The Query Tool (QT) enables users to browse studies, forms and CDEs, to select clinical data, use filters, and to sort and combine records. Using the GUID and a standard vocabulary via CDEs in forms, the QT provides an efficient means to reuse data by searching through volumes of aggregated research data across studies, find the right datasets to download and perform offline analysis using additional tools (e.g. SAS, SPSS, etc.). The statistical 'R-box' tool, integrated with BRICS, has been incorporated in the QT, to support analysis without having to download data.

The QT has several ways to search for data. By default, the user is presented with all studies in the data repository that have data submitted against them. Users can use the QT to search for desired data by searching by study, or across studies by form or an individual data element (Figure 3a).

Each column of data in a QT result represents a well-defined element in the Data Dictionary. Users can refine results by selecting from the list of allowed element permissible values, like male or female, or move sliders to select a range of numeric values, like age or outcome scores (shown in Figure 3b).

In addition to providing tools to aid data discovery, the QT supports interactive features that facilitate analysis and practical

Figure 3a. The Query Tool functionality is used to browse studies and forms, search data within forms and across studies. Example from the Parkinson's Disease Biomarker Program BRICS instance.

The screenshot shows the PDBP DMR Query Tool interface. At the top, there's a navigation bar with links like Home, Workspace, ProForms, GUID, Data Dictionary, Data Repository, Query, Meta Study, and Account Management. Below this, there's a 'Step 1 Filter Data' and 'Step 2 Refine Data' section. The 'Data Cart' on the left allows users to select forms to refine their query. The 'Filters' panel on the left shows a list of filters, including 'PDBP Adverse Events' and 'AdvsEvtDuringStudyInd'. The main area displays a 'Data Table View' of 'PDBP Adverse Events' with 4 rows of data. A blue arrow points from the 'AdvsEvtDuringStudyInd' filter to the 'AdvsEvtDuringStudyInd' column in the table.

GUID	AgeYrs	AdvsEvtDuringStudyInd	AdvsEvtStartTime	AdvsEvtSeverity	AdvsEvtRelatedness
PDAJ40URR	59	Yes	2013-12-12T00:00:00Z	Mild	Possible
PDFB758RMJ	61	Yes	2015-02-25T00:00:00Z	Mild	Possible
PDDC904HWE	72	Yes	2016-07-07T00:00:00Z	Mild	Definite
PDVH048WXT	68	Yes	2017-04-19T00:00:00Z	Mild	Definite

Figure 3b. The Query Tool can be utilized by users to select from a list of data elements that exist or are part of a form structure.

use of the data through attribute-based filtering capabilities, based on the data element type.

Various datasets (e.g. clinical, cognitive, demographic) are available within the repositories that are integrated to the BRICS instances. Data can be shared in CSV file format for download, and/or stored in the Meta Study module for further analysis, research, and reference.

Biomedical data management use case

The Parkinson's Disease Biomarker Program (PDBP) signifies the importance of Parkinson's disease biomarker discovery process, which requires data replication and validation prior to clinical trial use²⁸. Making both research data and workflow process findable, accessible, interoperable, and reusable was an important design consideration during the development of the PDBP system. The system consists of two major components - (a) Drupal-based portal, and (b) the PDBP Data Management Resource (DMR). The portal is publicly accessible to users for obtaining policy, stakeholders, individual PI, and specific study information, including summary data, and news (see [PDBP site](#)). The PDBP DMR is comprised of the previously discussed BRICS modules (shown in [Figure 2](#)) and incorporated the Parkinson's disease CDEs into its Data Dictionary²⁹. Use of CDEs results in making data FAIR by harmonization of clinical, imaging, genomics, laboratory, and biospecimen data. The CDEs are easily accessible from multiple open resources - the PDBP data dictionary³⁰, the NINDS CDE project¹⁵, and the NIH CDE repository²¹. The DMR is securely managed with capabilities for account verification, GUID generation, data submission, validation, workflows, access, and biospecimen data management. A GUID is generated for each subject on their initial visit and is attached to the deidentified data. The GUID makes data reusable by enabling the aggregation of all research data (clinical,

imaging, genomic, and biomarker) for a specific subject, both within a single study and across many PDBP studies.

The ProFoRMS module (shown in [Figure 2](#)) is used to schedule Parkinson's Disease subject visits and capture data (including the GUID) via a web-based assessment form tool. It provides capabilities for real time data entry and automatic data harmonization and ensures data quality assurance prior to storage within the PDBP repository. Each of the questions in the PDBP DMR assessment form is associated to a CDE that supports reusability and interoperability of PDB data^{28,31}. The ProFoRMS also provides automatic assignment of specific forms to individualized cohorts based on protocol design, and quality assessment of data prior to uploading to the PDBP Data Repository.

The authorized PDBP users can use the QT for accessing data across studies and aggregate data based on assessment forms and CDEs, allowing for the linkage of biosample data to demographics data. More complex queries can be created by linking clinical data from ProFoRMS with imaging data from the MIPAV module and with corresponding biospecimens/biosamples.

Data can be downloaded directly from the PDBP data repository and/or from the Query Tool to be analyzed by researchers using their preferred tools. Because the DMR database contains only de-identified data, all data uploaded to the DMR can be shared with the scientific community. Use of standard operating procedures has resulted in harmonization of biospecimens/biosamples with the DMR Biosample Order Manager tool, which enables linking clinical and biorepository data³². The PDBP data, queries and other metadata described for the research can be loaded into the Meta Study module and through

the Meta Study user interface researchers can generate DOIs that can be referenced in research articles.

Biomedical Program Application

The initial deployment of BRICS was to support the U.S. Department of Defense's (DoD) and the National Institute of Neurological Disorders and Strokes (NINDS), **FITBIR** project. The core functionalities for FITBIR were reusable for developing **PDBP**, as well other biomedical programs.

A few highlights of the data repositories resulting from the implementation of BRICS instance for the biomedical programs are provided below -

Federal Interagency Traumatic Brain Injury Research (FITBIR), is a BRICS instance developed to advance comparative effectiveness research in support of improved diagnosis and treatment for those who have sustained a TBI³³. The FITBIR repository stores data provided by TBI researchers and has accepted high quality research data from several studies, regardless of funding source and location. The DoD and NINDS provides funding for TBI human subject studies (both retrospective and prospective) and have required the research grantees to upload their clinical, imaging, and genomic data to FITBIR. As of 2018, there are 157 studies in FITBIR, spanning nearly hundred PIs, dozens of universities and research systems, the DoD, and the NIH. Data on 69,208 subjects, including more than 82,000 clinical image 3D data sets that are part of the repository. Currently, there are a total of 1,857,926 records in FITBIR. Data provided to FITBIR for broad research access are expected to be made available to all users within six months after the award period ends.

Parkinson's Disease Biomarkers Program Data Management Resource (PDBP DMR), is a BRICS instance developed to support new and existing research and is a resource for promoting biomarker discovery for Parkinson's disease funded by NINDS, NIH. At the center of the PDBP effort is its DMR. The PDBP DMR uses a system of standardized data elements and definitions, which makes it easy for researchers to compare data to previous studies, access images and other information, and order biosamples for their own research. PDBP's needs have accelerated BRICS system development, such as enhancements to the ProFoRMS data capture module, also with an investment into a BRICS plug-in for managing biosamples. The PDBP DMR now contains over 1,500 enrolled subjects, 1,415 of whom have biorepository samples. Also, PDBP has currently a total of 55,400 records.

eyeGENE, has a BRICS instance to support the National Ophthalmic Disease Genotyping and Phenotyping Network³⁴. It is a research venture created by the National Eye Institute (NEI) to advance studies of eye diseases and their genetic causes, by giving researchers access to DNA samples and clinical information. Data stored in eyeGENE is cross-mapped to Logical Observation Identifiers Names and Codes terminology (LOINC) interoperability data standards³⁵. Currently, eyeGene has 146,024 records with 6,400 enrolled subjects.

Informatics Core of Center for Neuroscience and Regenerative Medicine (CNRM) has a BRICS instance to support the CNRM medical research program with collaborative interactions between the U.S. DoD, NIH, and the Walter Reed National Military Medical Center. The Informatics Core provides services such as electronic data capture and reporting for clinical protocols, participation in national TBI research and data repository community, integration of CNRM technology requirements, and maintenance of a CNRM central data repository³⁶. In addition, the Informatics Core has played an important role in the development of multiple BRICS modules used by FITBIR.

Common Data Repository for Nursing Science (cdRNS) has a BRICS instance to support the National Institute of Nursing Research (NINR) mission - to promote and improve the health of individuals, families, and communities³⁷. To achieve this mission, NINR supports and conducts clinical and basic research and research training on health and illness. This research spans and integrates the behavioral and biological sciences, and that develops the scientific basis for clinical practice³⁸. The NINR is a leading supporter of clinical studies in symptom science and self-management research. To harmonize data collected from clinical studies, NINR is spearheading an effort to develop CDEs in nursing science. Currently, there are 1,358 records in the cdRNS instance of BRICS.

The Rare Diseases Registry, has a BRICS instance for the Rare Diseases Registry (RaDaR) program of the National Center for Advancing Translational Sciences (NCATS). It is designed to advance research for rare diseases³⁹. Because many rare diseases share biological pathways, analyses across diseases can speed the development of new therapeutics. The goal is to build a Web-based resource that integrates, secures, and stores de-identified patient information from many different registries for rare diseases, all in one place.

Discussion

The informatics system utilizes the Open Archival Information System (OAIS) model for preserving information for a designated community (group of potential consumers and multiple stakeholders). The implementation of the model highlights the importance of developing Submission, Archival, and Dissemination Information Packages (**Figure 2**, SIPs, AIPs and DIPs) for longer term data preservation and reuse⁴⁰. The primary producers of the data for the informatics system are the researchers and staff associated with each of the biomedical programs. Clinical data SIPs are produced for each of the instances by using eCRFs and imaging data SIPs are produced by the Image Submission tool. The CDEs and data dictionaries for the various BRICS instances support the development of archival information packages (AIPs), which are stored in distinct data repositories identified by the biomedical research programs⁴¹. The portability of the informatics software is also possible by recent deployment to the National Trauma Research (NTR) data repository development work⁴². In contrast to most of the centrally managed repositories within the NIH, the informatics software hosted for NTR is within a secure Amazon Web Services cloud platform. Deploying in the cloud

environment enhances data access, sharing, and reuse of biomedical research data at larger scale⁴³.

Supporting the FAIR principles

The FAIR (Findable (F), Accessible (A), Interoperable (I), and Reusable (R)) principles state that stewardship of digital data should promote discoverability and reuse of digital objects, which includes data, metadata, software and workflows⁴⁴. In addition, the principles posit that data and metadata should be accompanied by persistent identifiers (PIDs), indexed in a searchable resource, retrievable by their identifiers, and use vocabularies that meet domain relevant community standards. The principles serve as guidelines for developing systems that can improve data discovery and reuse. In Table 1, we have correlated the various BRICS functional components, which contribute towards making data FAIR for biomedical research programs.

Unique identification that is machine-resolvable with a commitment to persistence is fundamental for providing access

to data and metadata⁴⁵. In the context of the informatics system discussed here, GUID does not imply findability on the web, however, it purports findability of research participant data within a BRICS instance. Authorized researchers can use GUID to link together all submitted information for a single participant, even if data was collected at different locations and/or for different purpose(s).

Several identifier schemes (e.g. DOI, Handle system, Identifiers.org, Uniform Resource Identifiers) vary in their characteristics⁴⁶. A fundamental difference in an identifier scheme can be in the management (centrally or locally) of the resolver. For example, in the DOI scheme, a dereferencing service (e.g. Datacite or CrossRef) serves as a resolver that redirects the identifier to the actual content and the metadata.

The DOIs generated by BRICS is through the Interagency Data ID Service (IAD), which is operated by the U.S. Department of Energy Office of Scientific and Technical Information (OSTI).

Table 1. Informatics functional components that support the FAIR (Findable (F), Accessible (A), Interoperable (I), and Reusable (R)) principles. The FAIR principles listed in the table are from the cited reference ⁴⁴.

FAIR Principles	BRICS Functional Components					
	GUID	Data Dictionary	Data Repository	ProFoRMs	Query Tool	MetaStudy
Findable						
Data are assigned a globally unique and eternally persistent identifier	x	x	x			x
Data are described with rich metadata		x	x	x	x	x
Metadata clearly and explicitly include the identifier of the data it describes			x			x
(Meta) data are registered or indexed in a searchable resource		x	x			
Accessible						
(Meta) data are retrievable by their identifier using a standardized communications protocol		x				
The protocol is open, free, and universally implementable		x				
The protocol allows for an authentication and authorization procedure, where necessary	x	x	x	x	x	x
Metadata are accessible, even when the data are no longer available		x				
Interoperable						
(Meta) data use a formal, accessible, shared, and broadly applicable language for knowledge representation		x	x			
(Meta) data use vocabularies that follow FAIR principles		x	x	x		
(Meta) data include qualified references to other (meta) data		x			x	
Reusable						
(Meta) data have a plurality of accurate and relevant attributes		x	x		x	x
(Meta) data are released with a clear and accessible data usage license		x	x			x
(Meta) data are associated with their provenance		x	x			x
(Meta) data meet domain-relevant community standards		x	x			

The IAD service acts as a bridge to DataCite, which is one of the major registries of DOIs. The DOIs are assigned to individual research studies and are findable within the established repositories, available also from open sites with core metadata supported via Data Tag Suite (DATS) 2.2⁴⁷.

The availability of an automated validation tool with the informatics system makes CDE findable in the Data Dictionary and ensures for data quality and consistency. The system provides for an automated means of mapping CDEs to other informatics systems data dictionaries, e.g. CDISC⁴⁸. CDEs are made available through public websites (e.g. National Library of Medicine (NLM), NINDS CDE project, CDISC, etc.) to make data interoperable. Usability of data is enhanced by the adoption of standard imaging formats (e.g. DICOM, NIFTI, etc.). The informatics system also supports data discoverability across multiple repositories through the application of the biomedical and healthCare Data Discovery Index Ecosystem (bioCADDIE)⁴⁹.

Conclusion

Data confidentiality, integrity and accessibility are essential elements of responsible biomedical research data management. Community-wide data sharing requires development and application of informatics systems that promote collaboration and sustain data integrity of research studies within a secure environment. The informatics system presented above enables researchers to efficiently collect, validate, harmonize, and analyze research datasets for various biomedical programs. Integration of the CDE methodology with the informatics design results in sustainable digital biomedical repositories that ensure higher data quality. Aggregating data across projects, regardless of location and data collection time can define study populations of choice, for exploring new hypotheses based-research.

Data availability

Underlying data

All data underlying the results are available as part of the article and no additional source data are required

Software availability

Source code available from: <https://github.com/brics-dev/brics>

Archived source code at time of publication: <http://doi.org/10.5281/zenodo.3355727>⁵⁰

License: Other (open). Full license agreement is available from GitHub (<https://github.com/brics-dev/brics/blob/master/License.txt>)

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgment

The authors thank Mr. Denis von Kaeppler, Center for Information Technology, National Institutes of Health for helpful discussions and suggestions during the preparation of the manuscript, Ms. Abigail McAuliffe and Mr. William Gandler, Center for Information Technology, National Institutes of Health for editing the manuscript.

The opinions expressed in the paper are those of the authors and do not necessarily reflect the opinions of the National Institutes of Health.

References

- Sarkar IN: **Biomedical informatics and translational medicine**. *J Transl Med*. 2010; 8: 22.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Payne PR: **Chapter 1: Biomedical knowledge integration**. *PLoS Comput Biol*. Public Library of Science. 2012; 8(12): e1002826.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Harris PA, Taylor R, Thielke R, *et al.*: **Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support**. *J Biomed Inform*. 2009; 42(2): 377–381.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- for Clinical Translational Science UC: **REDCap: Research electronic data capture**. UIC Center for Clinical and Translational Science; 2018.
[Reference Source](#)
- Murphy S, Wilcox A: **Mission and Sustainability of Informatics for Integrating Biology and the Bedside (i2b2)**. *EGEMS (Wash DC)*. 2014; 2(2): 1074.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Scheufele E, Aronzon D, Coopersmith R, *et al.*: **tranSMART: An Open Source Knowledge Management and High Content Data Analytics Platform**. *AMIA Jt Summits Transl Sci Proc*. 2014; 2014: 96–101.
[PubMed Abstract](#) | [Free Full Text](#)
- Herzinger S, Gu W, Satagopam V, *et al.*: **SmartR: an open-source platform for interactive visual analytics for translational research data**. *Bioinformatics*. 2017; 33(14): 2229–2231.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Goff SA, Vaughn M, McKay S, *et al.*: **The iPlant Collaborative: Cyberinfrastructure for Plant Biology**. *Front Plant Sci*. 2011; 2: 34.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Devisetty UK, Kennedy K, Sarando P, *et al.*: **Bringing your tools to CyVerse Discovery Environment using Docker [version 1; peer review: 3 approved]**. *F1000Res*. 2016; 5: 1442.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Scott A, Courtney W, Wood D, *et al.*: **COINS: An Innovative Informatics and Neuroimaging Tool Suite Built for Large Heterogeneous Datasets**. *Front Neuroinform*. 2011; 5: 33.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Landis D, Courtney W, Dieringer C, *et al.*: **COINS Data Exchange: An open platform for compiling, curating, and disseminating neuroimaging data**. *Neuroimage*. 2016; 124(Pt B): 1084–1088.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mohr C, Friedrich A, Wojnar D, *et al.*: **qPortal: A platform for data-driven biomedical research**. *PLoS One*. 2018; 13(1): e0191603.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cimino JJ, Ayres EJ, Remennik L, *et al.*: **The National Institutes of Health's Biomedical Translational Research Information System (BTRIS): design, contents, functionality and experience to date**. *J Biomed Inform*. 2014; 52:

- 11–27.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Thompson HJ, Vavilala MS, Rivara FP: **Chapter 1 Common Data Elements and Federal Interagency Traumatic Brain Injury Research Informatics System for TBI Research.** *Annu Rev Nurs Res.* 2015; **33**: 1–11.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Silva J, Wittes R: **Role of clinical trials informatics in the NCI's cancer informatics infrastructure.** *Proc AMIA Symp.* 1999; 950–954.
[PubMed Abstract](#) | [Free Full Text](#)
16. **Common Data Element (CDE) - Clinfowiki.** [cited 3 Apr 2018].
[Reference Source](#)
17. **NINDS Common Data Elements.** [cited 3 Apr 2018].
[Reference Source](#)
18. Rubinstein YR, McInnes P: **NIH/NCATS/GRDR® Common Data Elements: A leading force for standardized data collection.** *Contemp Clin Trials.* 2015; **42**: 78–80.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Moore SM, Schiffman R, Waldrop-Valverde D, *et al.*: **Recommendations of Common Data Elements to Advance the Science of Self-Management of Chronic Conditions.** *J Nurs Scholarsh.* 2016; **48**(5): 437–447.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Sheehan J, Hirschfeld S, Foster E, *et al.*: **Improving the value of clinical research through the use of Common Data Elements.** *Clin Trials.* 2016; **13**(6): 671–676.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. **Glossary.** U.S. National Library of Medicine: 2012.
[Reference Source](#)
22. Hall D, Huerta MF, McAuliffe MJ, *et al.*: **Sharing heterogeneous data: the national database for autism research.** *Neuroinformatics.* 2012; **10**(4): 331–339.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Haak D, Page CE, Deserno TM: **A Survey of DICOM Viewer Software to Integrate Clinical Research and Medical Imaging.** *J Digit Imaging.* 2016; **29**(2): 206–215.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Shah J: **Medical Image Processing, Analysis and Visualization.** [cited 6 Nov 2017].
[Reference Source](#)
25. O'Reilly PD: **Federal Information Security Management Act (FISMA) Implementation Project.** 2009.
[Reference Source](#)
26. National Institute of Standards, Technology: **FIPS 200, Minimum Security Requirements for Federal Info and Info Systems | CSRC.** [cited 7 Feb 2018].
[Reference Source](#)
27. Nist SP: **800-53, Revision 3. Recommended Security Controls for Federal Information Systems and Organizations.** 2009; 28–29.
[Reference Source](#)
28. Gwinn K, David KK, Swanson-Fischer C, *et al.*: **Parkinson's disease biomarkers: perspective from the NINDS Parkinson's Disease Biomarkers Program.** *Biomark Med.* 2017; **11**(6): 451–473.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Grinnon ST, Miller K, Marler JR, *et al.*: **National Institute of Neurological Disorders and Stroke Common Data Element Project - approach and methods.** *Clin Trials.* 2012; **9**(3): 322–329.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. PDBP: **Parkinson's Disease Biomarkers Program | PDBP: Parkinson's Disease Biomarkers Program.**
[Reference Source](#)
31. Rosenthal LS, Drake D, Alcalay RN, *et al.*: **The NINDS Parkinson's disease biomarkers program.** *Mov Disord.* 2016; **31**(6): 915–923.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. **How To Guide | PDBP.** [cited 20 Dec 2018].
[Reference Source](#)
33. Index | FITBIR: **Federal Interagency Traumatic Brain Injury Research Informatics System.** [cited 6 Nov 2017].
[Reference Source](#)
34. **eyegene.nih.gov.** [cited 6 Nov 2017].
[Reference Source](#)
35. **LOINC — The freely available standard for identifying health measurements, observations, and documents.** [cited 6 Nov 2017].
[Reference Source](#)
36. **CNRM Data Repository.** [cited 6 Nov 2017].
[Reference Source](#)
37. **cdRNS.** [cited 6 Nov 2017].
[Reference Source](#)
38. **Mission & Strategic Plan | National Institute of Nursing Research.** [cited 6 Nov 2017].
[Reference Source](#)
39. **Rare Diseases Registry Program (RaDaR): National Center for Advancing Translational Sciences.** 2017. [cited 6 Nov 2017].
[Reference Source](#)
40. Navale V, McAuliffe M: **Long-term preservation of biomedical research data [version 1; peer review: 4 approved, 1 approved with reservations].** *F1000Res.* 2018; **7**: 1353.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
41. Navale V, Ji M, McCreedy E, *et al.*: **Standardized Informatics Computing Platform for Advancing Biomedical Discovery Through Data Sharing.** *bioRxiv.* 2018; **259465.**
[Publisher Full Text](#)
42. Price MA, Bixby PJ, Phillips MJ, *et al.*: **Launch of the National Trauma Research Repository coincides with new data sharing requirements.** *Trauma Surg Acute Care Open.* 2018; **3**(1): e000193.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
43. Navale V, Bourne PE: **Cloud computing applications for biomedical science: A perspective.** *PLoS Comput Biol.* 2018; **14**(6): e1006144.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
44. Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al.*: **The FAIR Guiding Principles for scientific data management and stewardship.** *Sci Data.* 2016; **3**: 160018.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
45. Starr J, Castro E, Crosas M, *et al.*: **Achieving human and machine accessibility of cited data in scholarly publications.** *PeerJ Comput Sci.* 2015; **1**: pii: e1.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
46. Guralnick RP, Cellinese N, Deck J, *et al.*: **Community next steps for making globally unique identifiers work for biocollections data.** *Zookeys.* 2015; (494): 133–154.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
47. Sansone SA, Gonzalez-Beltran A, Rocca-Serra P, *et al.*: **DATS, the data tag suite to enable discoverability of datasets.** *Sci Data.* 2017; **4**: 170059.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
48. Park YR: **CDISC Transformer: a metadata-based transformation tool for clinical trial and research data into CDISC standards.** *KSII TIS.* 2011; **5.**
[Publisher Full Text](#)
49. Ohno-Machado L, Sansone SA, Alter G, *et al.*: **Finding useful data across multiple biomedical data repositories using DataMed.** *Nat Genet.* 2017; **49**(6): 816–819.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
50. **brics-dev: brics-dev/brics: Iron man (Version v1.0.0).** *Zenodo.* 2019.
<http://www.doi.org/10.5281/zenodo.3355727>

Open Peer Review

Current Peer Review Status: ? ✖

Version 1

Reviewer Report 03 December 2019

<https://doi.org/10.5256/f1000research.20997.r55222>

© 2019 Clark T. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Timothy W. Clark

School of Medicine and School of Data Science, University of Virginia, Charlottesville, VA, USA

This is an article which one eventually learns is about the BRICS system, developed for NINDS by software consultants at General Dynamics and Sapient. It is poorly organized, and seems to take little care to actually explain the system's motivation and use in a coherent fashion understandable to the reader.

The authors begin with an overgeneralized and basically non-informative abstract and introduction which make a number of broad statements about translational research and its requirements for enabling informatics. But one does not discover the actual real focused topic of the article until far down in the text.

One does not learn, for example, until page 8, the specific motivation of the system:

"The initial deployment of BRICS was to support the U.S. Department of Defense's (DoD) and the National Institute of Neurological Disorders and Strokes (NINDS), FITBIR project. The core functionalities for FITBIR were reusable for developing PDBP, as well other biomedical programs."

Obviously there has been a major software development effort conducted by contractors funded and managed by NINDS, and this article is attempting to report on what was accomplished. The implementation is not interesting in itself, using mostly 20-year-old technology. However, the scale of the software effort and its importance to researchers working on certain NIH funded projects, would seem to require something far better as a report out, whether or not it was accepted by a peer-reviewed journal.

Crucial rationales for technical decisions made, and contexts in which the system is used, are omitted or poorly motivated, or described inaccurately. For example, we learn on page 4 that:

"The Data Layer consists of open source databases such as PostgreSQL, Virtuoso databases, file

servers, and data persistence frameworks."

Virtuoso is not an open-source database. It is an enterprise-class RDF graph store. Nowhere up to now has the need for an RDF graph store been explained, and it is not touched on again until page 6, where we learn that:

"The raw data is processed through the 'NextGen Connect' tool (integrated interface engine) and Resource Description Framework (RDF) data interchange tool."

And that is all we ever learn about the use of RDF or any related semantic technologies in this system. Are there OWL Ontologies involved? Why was the decision made to use them and which ones were selected? Why is Virtuoso used in conjunction with PostgreSQL relational store? What makes this combination necessary? We hear nothing of this.

This reviewer prepared a line-by-line discussion of the text which can be found [here](#). Suffice it to say that there are many significant issues with this article - issues of poor exposition, lack of required detail or context, imprecision, or simply misleading statements.

Another example:

"The Meta Study module is used for meta-analysis of the data as well as a collaboration tool between scientific groups. "

That sentence is all we ever hear of the ability to perform meta-analysis, or any of the challenges it poses.

Or this:

"Deploying in the cloud environment enhances data access, sharing, and reuse of biomedical research data at larger scale".

In fact the reason to deploy things in the cloud is for rapid horizontal and vertical scaling. It has nothing to do with data sharing and reuse. The authors cite an article by one of them (Navale & Bourne 2018), to back up their incorrect claim - but that article directly contradicts this claim.

I strongly recommend to the authors that they engage a high-quality technical writing firm competent in bioinformatics to revise the text, paying close attention to precision, providing needed context for technical choices, context of usage and overall motivation of the system, and in general, thoughtful informative exposition.

Is the rationale for developing the new method (or application) clearly explained?

No

Is the description of the method technically sound?

Partly

Are sufficient details provided to allow replication of the method development and its use

by others?

No

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

No source data required

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Biomedical informatics. Semantic technologies. Cloud computing frameworks. Neuroscience. Data science.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 07 Jul 2020

Vivek Navale, National Institutes of Health, USA, Bethesda, USA

1. Thank you for the comments. We have revised the manuscript extensively and addressed the detailed comments that you provided us. The organization of the manuscript has been made easier for readers to follow, appropriate headings have been contextualized throughout the paper. The introduction has been refocused, motivation highlighted and the specific points that were brought to our attention have been addressed in the manuscript.
2. We have revised the abstract and focused on the BRICS functional components, services deployed for research data life cycle management and demonstrated the application to various biomedical research programs. The common data element concept has been contextualized and the significance to this work has been highlighted.
3. Thank you for the comments on the BRICS Architecture section. We have made major revisions to this section to explain the design choices during the development work, lessons learned and accurately depicted the current architecture in Figure 1.
4. We have clarified in the Data Submission section that institutional grants (e.g. DOD, NIH) support disease specific research and mandate data to be submitted to a specific BRICS instance.
5. We have revised the section to state that after data has been validated and uploaded to the repository, an original copy of the user submitted data (raw data) is maintained in the repository that can be accessed by the Query Tool.

6. Thank you for your suggestions. We have revised the section to clarify that data quality and consistency of submissions is enhanced by validation, using domain specific Data Dictionaries.

7. We agree that the biocaddie.org is not available, the project has been discontinued, hence access to BRICS repositories via bioCADDIE will not be possible. Therefore, we have removed the statements from the revised manuscript.

8. We have revised the section on BRICS instances and specified that the National Trauma Research Repository utilizes Cloud Computing for providing infrastructure as-a-service.

Competing Interests: I have no competing interests to disclose.

Reviewer Report 18 November 2019

<https://doi.org/10.5256/f1000research.20997.r56129>

© 2019 Kim H. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hyeoneui Kim

School of Nursing, Duke University, Durham, NC, USA

This paper introduces the Biomedical Research Informatics Computing System (BRICS), a comprehensive platform that supports researchers collect, store, analyze, and securely share research data. The underlying principle that motivated and enabled the implementation of this system is the FAIR (Findable, Accessible, Interoperable, and Reusable) principle for biomedical data.

The authors provided clear and highly informative descriptions of the architecture and the approaches to implementing the key functional components. The related initiatives and programs that aim at improving data use and reuse and the gaps found in them provide a convincing context for BRICS development. The figures presented in the paper adequately describe the structure, functions, and workflows of BRICS. The research programs and initiatives that already utilize BRICS introduced in the paper are strong evidence that supports the BRICS approach. All in all, this is a well-written, very informative paper. The changes suggested below could help strengthen this paper even more:

1. The conventional section structure that includes methods and results might not fit well with this paper. If F1000Research allows some flexibility in the manuscript structure, it will help readers follow the progress of the content by changing the Method section to something like BRICS Functionalities and Components (or something along this line) and Result to BRICS Instance (or use cases). Also, the query module explained in the result section can be included in the BRICS functionalities and components section.

2. A brief description of how the study level metadata (e.g., sample size, study design, study location, etc.) are captured would be helpful as study-level metadata are among the most frequently used parameters for data/dataset search.
3. Accessing a broader scope of data is one of the main motivations that researchers would adopt this type of data platform. Therefore, although it is not related to the technological development of BRICS, introducing more information on the data sharing policies (i.e., data use agreement and DAC's responsibilities) would be beneficial.
4. Please correct minor errors, such as typos and introducing BRICS first without fully spelling out the acronym.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: biomedical informatics, standardized data representation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 07 Jul 2020

Vivek Navale, National Institutes of Health, USA, Bethesda, USA

1. Thank you for the suggestions. To improve clarity, we have redefined method section as BRICS System Design and Architecture, followed by two sections - Data Submission and Processing, Sharing and Access sections. We have integrated the Query Tool description within the section the Data Access section.

2. Under the data submission and processing section, we have added information on study level metadata, that is entered manually through a graphical user interface, when a BRICS instance is used. The examples of Metadata fields include title, organization, PI, data, funding source and ID's, study type(s), and keywords that enable users to search for detailed information (e.g. clinical trial Grant ID(s), start and end dates for grants, therapeutic agents, sample size, publications, and forms used).

We have indicated that each of the BRICS instance exposes metadata and summary consistent with their respective program goals. We have provided an example, FITBIR provides a metadata visualization tool that graphically supports searching study identification (shown here <https://fitbir.nih.gov/visualization>).

3. Thank you for the suggestion. We have added information in the Data Sharing and Access section indicating that each instance of BRICS supports the data sharing policies consistent with their respective program. Research data is maintained in a private state until a year after the grant end date, and after that time, data is moved to a shared state where all users with approval from DAC can have access to the data. The DAC is comprised of government program officials responsible for each of the BRICS instances, who evaluate the data access requests and approve or disapprove the request. A detailed information for each of the BRICS instances can be gleaned from the site information (web site links) provided under the BRICS instance section.

Competing Interests: I have no competing interests for my comments.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research