Journal of Data, Information and Management (2021) 3:21–27
https://doi.org/10.1007/s42488-020-00040-0

ORIGINAL ARTICLE

An overview of biomedical platforms for managing research data

Vivek Navale[1] · Denis von Kaeppler[1] · Matthew McAuliffe[1]

Received: 10 November 2020 / Revised: 17 December 2020 / Accepted: 28 December 2020 / Published online: 23 January 2021

Abstract

Biomedical platforms provide the hardware and software to securely ingest, process, validate, curate, store, and share data. Many large-scale biomedical platforms use secure cloud computing technology for analyzing, integrating, and storing phenotypic, clinical, and genomic data. Several web-based platforms are available for researchers to access services and tools for biomedical research. The use of bio-containers can facilitate the integration of bioinformatics software with various data analysis pipelines. Adoption of Common Data Models, Common Data Elements, and Ontologies can increase the likelihood of data reuse. Managing biomedical Big Data will require the development of strategies that can efficiently leverage public cloud computing resources. The use of the research community developed standards for data collection can foster the development of machine learning methods for data processing and analysis. Increasingly platforms will need to support the integration of data from multiple disease area research.

Keywords Biomedical platforms · Data processing · Sharing · Reuse · Data management

1 Introduction

Biological data arises from a variety of sources – genomic sequencing, imaging studies, clinical, phenotypic, ecological, and microscopic research work. Harnessing the power of data requires it to be findable, accessible, interoperable, and reusable (FAIR), (Wilkinson et al. 2016). Large scale initiatives like the All Of Us Research Program (AOU) need platforms to support multi-modal data integration, modeling, and linking of data from different sources. Managing the research data life cycle requires biomedical platforms to support comprehensive data management plans (Griffin et al. 2017).

Biomedical data platforms can provide scalable infrastructures (hardware and software), secure services to ingest, process, validate, curate, store, and share data. These platforms support workflow(s), data analyses, visualization tools, and access to storage repositories. Research communities need general-purpose biological, clinical, translational, and disease area research platforms.

✉ Vivek Navale
Vivek.Navale@nih.gov

[1] Center for Information Technology, National Institutes of Health, 9000 Rockville Pike, Bldg 12A, (Rm 4041), Bethesda, MD 20892, USA

Several national and international platforms have been developed to support biological research. ELIXIR supports life science researchers from 23 European countries (https://elixir-europe.org/platforms) and is coordinated by the European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI), the de.NBI provides bioinformatics services to life science researchers in Germany (Tauch and Al-Dilaimi 2019), and the EMBL Australia Bioinformatics Resource coordinates life sciences and biomedical researchers in Australia, (Schneider et al. 2017). A distributed biotechnology information network comprising over a hundred centers is supported by the Indian government (Krishnaswamy and Madhan Mohan 2016). Within the United States (US), there are several biological and biomedical research platforms. For example, the US National Science Foundation funded the CyVerse platform to enhance interdisciplinary collaborations in life sciences (Goff et al. 2011), (Merchant et al. 2016). The tranSMART supports clinical and translational research (Herzinger et al. 2017), the National Database for Autism (NDA) provides access to clinical, behavioral assessments and health outcomes from novel interventions for Autism research (Payakachat, Tilford, and Ungar 2016), the Federal Interagency Traumatic Brain Injury Research (FITBIR) supports traumatic brain injury research (https://fitbir.nih.gov/) and the

Global Alzheimer's Association Interactive Network (GAAIN) supports Alzheimer research activities (Toga 2017).

A comprehensive review of the various platforms used in biomedical research is not within the scope of this article. This article serves as an overview of platform capabilities, and the examples provided illustrate the diverse data content, infrastructure, services, tools, and methods for increasing access and use of biomedical research data.

1.1 Platform infrastructure

The examples shown in Table 1 are used for general-purpose life sciences, clinical and translational research, and disease area studies. The infrastructure for the various platforms can be distributed or centralized. ELIXIR, for example, is a distributed platform across national boundaries with software tools, computing, and training resources for life sciences research. The EMBL-EBL serves as a coordinating center (hub) connected to various centers of excellence (nodes). The de.NBI, a node of the ELIXIR platform has eight service centers with informatics capabilities for biomedical research, microbial research, proteomics, RNA analysis, standards-based systems biology, and tools for omics data and imaging, including reference database services.

Centralized platforms like tranSMART enable data from different sources to be integrated for data analysis, hypothesis generation, and cohort discovery for clinical research (Scheufele et al. 2014). It utilizes the Integrated Biology and Bedside (i2b2) system, which provides software tools for the collection and validation of clinical research data (Murphy and Wilcox 2014). The European Translational Informational and Knowledge Management Service (eTRIKS) platform is interfaced with tranSMART; eTRIKS provides analysis and visualization of omics, preclinical laboratory data, and clinical information (Bussery et al. 2018). The Collaborative Informatics and Neuroimaging Suite (COINS) developed by the Mind Research Network as an open-source centralized platform hosts and provides services for the compilation, curation, and dissemination of neuroimaging data from more than 18 sites spread throughout the US (Landis et al. 2016).

The Genomic Data Commons (GDC), supported by the US National Cancer Institute serves as a centralized repository for cancer genomics and associated clinical data. The GDC platform provides analytic pipelines to align raw sequencing data to the human genome, and identify mutations, copy-number alterations, and gene-expression changes. (Grossman et al. 2016)

Many large-scale biomedical platforms use cloud resources as part of their infrastructure e.g. the CyVerse, NDA, GDC (Table 1). For Big Data biomedical problems, cloud computing provides an environment for data and analytics sharing that can increase the likelihood of data reproducibility and reuse (Navale and Bourne 2018). Currently, in partnership with Google and Amazon Web Services, the NIH Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) program offers cloud resources (tools, services, training) to NIH researchers and grantees (https://cloud.cit.nih.gov/). Commercial biomedical platforms (e.g., DNAnexus) also provide a secure cloud computing environment for analyzing and integrating phenotypic, clinical, and genomic data (https://www.dnanexus.com/). Within Europe, a trusted cloud-based digital platform, the European Open Science Cloud (EOSC) is being developed to provide access to data and services across geographical borders and scientific disciplines (https://ec.europa.eu/info/research-andinnovation/strategy/goals-research-and-innovation-policy/open-science/european-open-science-cloud-eosc_en).

1.2 Data processing applications and services

A Web-based platform, e.g. the Galaxy (started in 2005) has enabled access to genomics, proteomics, metabolomics, and imaging data sets, and currently hosts more than 5, 500 tools as part of the Galaxy ToolShed (Afgan et al. 2018). Data from the European Genome-phenome Archive (EGA) can be downloaded to Galaxy by using the tools available in the Galaxy ToolShed (Hoogstrate et al. 2016), and systems have been designed to link tranSMART, Galaxy, and EGA for reusing human translational data (Zhang et al. 2017).

ELIXIR supports the use of biotools and biocontainers (Table 1). A comprehensive registry of software and databases can be accessed via the bio.tools registry; it enables researchers to find, understand, utilize, and cite the resources for their research. The resources include both simple command-line tools and online services, as well as databases and complex, multi-functional analysis workflows. The Biocontainers are computationally executable environments that are platform-independent, which can be used for installing bioinformatics software, and combining tools for implementing data analysis pipelines (da Veiga Leprevost et al. 2017). A set of guidelines have been developed by the BioContainers Community to make the bioinformatics software more discoverable, reusable and transparent (Gruening et al. 2018).

Biomedical Research Informatics Computing System (BRICS) supports programs in several disease areas, e.g. Traumatic Brain Injury and Parkinson's biomarker research (Table 1). To manage data from different sources (e.g., biosamples, clinical, omics, imaging data) the BRICS provides services for electronic data capture, access to data dictionaries, processing, and storage within disease-specific digital repositories (Navale et al. 2019). Other web-based tools such as the Research Electronic Data Capture (REDCap) are used for collecting and processing clinical data (Harvey 2018), and an integrated platform (qPortal) can be used for

Table 1 Summary of platform types with some highlights

Platforms	User Community	Content example	Infrastructure	Services and Tools	Links
General purpose					
ELIXIR	Bioinformaticians and life science researchers from 23 European countries.	Chemical biology, enzymes, interactions and pathways, evolution and phylogeny, genes and genome, proteins and proteomes, molecular and cellular structures.	Hub and node(s) model supported by cloud computing resources, authentication and authorization.	Bio.tools, Biocontainers, UseGalaxy.eu, OpenEBench, EDAM ontology, biotools Schema.	https://elixir-europe.org/platforms
EMBL-ABR	Life science researchers, biomedical and bioinformatics communities in Australia.	Large life sciences and heterogeneous research datasets.	Hub and 10 nodes, federated bioinformatics network across Australia	EMBL-ABR Tool registry, ToolsAU, (Biotools.org and STM powered by the ELIXIR Tools and Data registry).	https://www.emblaustralia.org/
de.NBI	Basic and applied life sciences researchers.	Human, plant and microbial research, genomic sequencing, transcriptomics, proteomics data.	Supported by the de.NBI cloud.	Eight service centers (Bigi, bioinfra.port, BioData, RBC, HD-hub, GCBN, de.NBI-Sysbio, CIBI).	https://www.denbi.de/
CyVerse	Life science researchers.	Microbial, plant, ecological, biomedical, sequencing and imaging data.	High performance computing, cloud computing, i-RODS, single sign on.	Discovery environment, Cyberduck, SciApps, Science APIs, BisQue Image Analysis, DNA Subway.	https://www.cyverse.org/
Platforms	User Community	Content example	Infrastructure	Services and Tools	Links
COINS	Neuroscience and biomedical researchers.	Neuroimaging data, clinical and other assessments data on human subjects.	Open source platform hosted by the Mind Research Network.	Assessment Manager, Query Builder, Participant Portal, COINS Data Exchange for data sharing.	https://coins.trendscenter.org/
BTISNET	Teachers, scientists and students.	Genome, Gene expression, Nucleotide Sequence, Protein Structure.	Distributed network of centers, supported by supercomputing and bioinformatics and interactive graphics facilities.	Open-source databases and tools for protein, genome analysis, protein structure prediction and drug design.	http://btisnet.gov.in/
Clinical and translational research					
tranSMART	Clinicians, researchers, pharmaceutical companies, patient advocacy groups, universities and governments.	Clinical, translational and genomics data. Patient/Study level clinical data (e.g. diagnosis, medications, lab results), subject level high dimensional data (e.g. gene and protein expression arrays).	Open-source, modular, web-application framework, ontology-driven architecture, utilizes components of i2b2 for queries, exploration & analysis.	DBMI Data Portal, search engine for real-time indexing, Dataset Explorer for analysis. ETL, syntactic and semantic mapping, community developed plugins, APIs and user interfaces.	https://www.i2b2transmart.org/about-us/
eTRIKS	Biomedical and translational researchers.	Preclinical, clinical, multi-Omics. Several biomedical projects - ABIRISK, OncoTrack	Open-Stack cloud-based platform. Bioaster, curated public biomedical studies.	Authentication, authorization and access audit services.	https://www.etriks.org/
Platforms	User Community	Content example	Infrastructure	Services and Tools	Links
Disease focused					
NDA	Autism researchers.	Data definitions for 800 autism measures from clinical, imaging and genomic research data. Individual-level clinical assessments, outcomes from novel interventions, treatments for children with autism.	Supported by a controlled-access data repository for autism research, Neuroinformatics Tool and Resources cloud (NITRC).	GUID, upload, download, validation, NITRC computational environment and the Laboratory of Neuro Imaging (LONI) Pipeline resources.	https://nda.nih.gov/
FITBIR	Traumatic brain injury researchers.	Phenotypic, imaging, genomic, biomarker data, demographic and outcome assessments for humans. Uses Common Data Elements.	Supported by BRICS. Hosted within the NIH Data Center.	GUID, Data dictionary, account management, query tool, protocol and form research management system, meta study, repository manager.	https://fitbir.nih.gov/
PDBP	Parkinson's disease researchers.	Clinical, neuroimaging, biofluid data. Uses Common Data Elements.	Supported by BRICS. Hosted within the NIH Data center, associated PDBP biorepository and Data management resource.	GUID, Data dictionary, account management, query tool, protocol and form research management system, meta study, repository manager.	https://pdbp.ninds.nih.gov/

Table 1 (continued)

Platforms	User Community	Content example	Infrastructure	Services and Tools	Links
Platforms	User Community	Content example	Infrastructure	Services and Tools	Links
GAAIN	Alzheimer's disease, dementia and researchers on aging	Clinical, imaging, genetic, proteomic data, cohort discovery and data exploration.	Federated, global network of data partners. Data cached in computer memory but never written to disk.	Three interfaces: GAAIN Scoreboard, Interrogator, Cohort Scout search through thousands of data attributes. Data exported into CSV files and mapped into the GAAIN schema using CDISC conventions.	http://www.gaain.org/
GDC	Cancer research community.	Genomic and clinical data from different sources, (e.g. TCGA, TARGET, International Cancer Genome consortium).	Hosted by the University of Chicago Data center. Data portal web interface to browse, query, and download data.	New cancer genomic data uploaded by Open GDC, Data Harmonization. Mapping of sequence data to current genome and transcriptome build provided	https://portal.gdc.cancer.gov/

CDISC - Clinical Data Interchange Standards Consortium, i-RODS - Open-source Data Management Software, DBMI - Department of BioMedical Informatics, STM - Search for Training Material, ETL - Extract, Transform, Load, TARGET - Therapeutically Applicable Research to Generate Effective Treatments, GUID - Globally Unique Identifier, TCGA - The Cancer Genome Atlas, i2b2 - Informatics for Integrating Biology & the Bedside, BRICS - Biomedical Research Informatics Computing System

the quantitative management of laboratory biological data (Mohr et al. 2018).

Metadata creation and description during biomedical research work is an important part of data processing. The Center for Expanded Data Annotation and Retrieval (CEDAR) system provides capabilities to assemble composite templates, using metadata acquisition forms when acquiring a biomedical dataset (Musen et al. 2015).

Data privacy is an important aspect to consider during data processing. Access to research with patient's personal information requires an institutional review board (IRB) approval. To mitigate the time required for IRB review and approval and to accelerate data reusability, risk-aware access control methods for processing needs should be considered early during the research planning phase of the work (Badji and Dankar 2018).

Ethical issues (e.g. consent and sharing patient data) related to the openness of the data will also need to be evaluated. To support various users, a research data warehouse framework can consist of segregated identified and de-identified clinical data repositories, with access protocols and governance rules for data processing (Danciu et al. 2014).

1.3 Enabling data access and reuse

Biomedical platforms have resources for tools and services that facilitate data access and reuse. ELIXIR has identified 19 Core Data Resources (CDR) that provides a wide range of capabilities, which includes access to data from high throughput functional genomic experiments, information on human protein-coding genes, comprehensive high-quality datasets related to rare diseases (orphan data), mass spectrometry-based proteomics data, protein sequencing data and other resources (Drysdale et al. 2020). Several of the CDRs support the use of bioschemas and the Schema.org markup in their websites to enhance the findability of research data (https://bioschemas.org/).

Ontologies (commonly controlled vocabularies) are useful to standardize the collection, description, querying, and interpretation of data. The Open Biological and Biomedical Ontology (OBO) Foundry promotes the usage of a set of principles in the development of ontologies, ontology models, such as the Gene Ontology (Smith et al. 2007). Various types of biological data use different ontologies, and selecting a bio-ontology requires knowledge about the specific domain with an understanding of biological systems (Malone et al. 2016). An online collaborative tool (e.g., OntoBrowser) can be helpful to map reported terms to a preferred ontology (i.e., code list) for data integration purposes (Ravagli, Pognan, and Marc 2017). Platforms such as tranSMart provide an ontology-based approach to map collected data to institution-specific or industry-standard formats.

Common Data Elements (CDEs) are used in clinical research studies, and represent a combination of the precisely defined questions (variable) that can be associated with a specified value ("Common Data Element (CDE) - Clinfowiki" n.d.). Data aggregation, meta-analyses, and cross-study comparisons are benefited by the use of CDEs (Sheehan et al. 2016). The BRICS platform supports the use of CDE methodology during data collection, it utilizes data dictionaries that are based on CDEs for specific disease areas, examples in Table 1 are the TBI and PDBP platforms (Navale et al. 2019).

Common Data Models (CDMs) are used to standardize the collection of research data, which can facilitate the aggregation and sharing of data. There are several CDMs available for specific uses, examples include the National Patient-Centered Clinical Research Network (PCORnet) and the Observational Medical Outcomes Partnership (OMOP). An evaluation of CDM use for longitudinal Electronic Health Record (EHR) based studies showed that the OMOP CDM best met the criteria for supporting data sharing (Garza et al. 2016). The AOU program is standardizing the EHR data by using the OMOP CDM. Methods to harmonize data from the i2b2 system to the OMOP model have also been provided recently (Klann et al. 2019).

Many standards and databases are available for data, metadata collection, and storage within databases and repositories. The Fairsharing.org provides a service to relate data and metadata standards with databases and data policies (Sansone et al. 2019). The 'FAIR' cookbook, provides 'recipes' for making different types of life science data FAIR (https://fairplus-project.eu/). For genomic data sharing the Global Alliance for Genomics and Health (GA4GH) an international consortium provides standards for responsibly collecting, storing, analyzing, and sharing genomic data (https://www.ga4gh.org/).

1.4 Managing biomedical research data

Advancements in genomic sequencing capabilities coupled with decreasing service costs have significantly increased (several hundred terabytes and more) the generation of genomic data within biomedical institutions. Detecting disease-causing genes requires a series of computing steps that eventually yield specific information useful in the clinical care of patients. Most of the raw genomic data from sequencing are less utilized after the completion of the analysis and is maintained within the core facilities and/or associated repositories.

Biomedical platforms will require the expansion of high-performance computing capabilities and storage repositories to meet the needs of many biological Big Data projects . These needs can impose budgetary challenges for even well-funded institutions. A strategy to consider by institutions is to migrate raw genomic data with infrequent access requirements to lower-cost cloud storage options. Cloud storage backup strategy can be extended for maintaining critical biomedical data in public clouds and to support disaster recovery plans.

The cloud model promotes the 'data at rest' concept, that is data can be produced, managed, and accessed at one location without having to download to individual user computers. Project-specific scalability of services can be implemented in a cloud-based platform with service cost incurred on usage. This approach can alleviate the need for periodic upgrade and refresh of computing and storage capacities thereby reducing the institutional IT budget costs from Big Data projects. It should be noted that the inherent advantage of using public clouds has some trade-offs, for example, data ingress (moving data to a public cloud) does not incur a cost, however, egress (moving data out) costs from a vendor cloud can be significant, especially if large scale data migrations are necessary at any time. There is a risk for data to be siloed by increasingly relying on a specific public cloud provider. To mitigate these risks, developing cloud interoperability strategies and methods (e.g. common application programming interfaces) that can enable communication between applications and services for two or more public cloud instances is a near-term need. The availability of interoperable cloud computing technologies can facilitate the portability of biomedical data between various cloud deployments.

Big Data brings other challenges as well – for example, there can be a significant time gap between initial data generation and the subsequent processing and analysis work needed to produce meaningful information for a large-scale project. Consistent use of data elements, models, dictionaries, and standard vocabularies during the data collection phase can mitigate manual, often laborious data curation that is time-intensive and expensive. The use of standardized methods can also facilitate machine readability, promote automation, and enhance the use of machine learning (ML) technologies for Big Data biomedical research.

Currently, ML methods are being utilized in medical imaging analyses for aiding in disease confirmation. Increasing development and application of ML methods for disease detection, prevention and prediction will require biomedical platforms to support the growth of ML-based data sets.

In our overview we gleaned that the current landscape of platforms supports some aspect of the FAIR principles, enhancing platform capabilities to satisfy all of the recommended principles will result in more effective data stewardship and management. The use of community-agreed data formats, metadata standards, tools, and services can improve data integration capabilities. We anticipate that new insights in disease area research will require platforms to support the integration of multiple data types - genomics, proteomics, imaging, phenotypic and clinical data for research projects. New modalities will also be required to address the challenge of determining commonalities across various disease area research.

2 Conclusion

Biomedical platforms are required for the collection, processing, analysis, storage, and access to research data. They can vary in scope and size from being general-purpose to disease-specific in nature. Increasingly cloud computing technology is being integrated with the platform architecture, to support Big Data projects. Several online software applications, methods, and services are available for researchers to use for their project needs. The diversity in standards, models, ontologies that can be used for managing research data requires both subject matter expertise and engagement with a discipline-specific research community. Overall, platforms will add value to biomedical research by supporting data to be FAIR. New insights in disease area research will require platforms to support the integration of clinical, imaging, phenotypic, genomic, and proteomic data that can contribute towards personalized and precise medical diagnosis and care.

Acknowledgements We thank Ms. Alicia A. Livinski, National Institutes of Health, for editing the manuscript.

Compliance with ethical standards

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

Conflicts of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

Afgan E, Baker D, Batut B, van den Beek M, Bouvier M, Cech M, Chilton J et al (2018) The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46(W1):W537-44

Badji R, Dankar FK (2018) A risk-aware access control model for biomedical research platforms. Proceedings of the 4th International Conference on Information Systems Security and Privacy. https://doi.org/10.5220/0006608403220328

Bussery J, Denis L-A, Guillon B, Liu P, Marchetti G (2018) eTRIKS platform: conception and operation of a highly scalable cloud-based platform for translational research and applications development. Comput Biol Med 95(April):99–106

Common Data Element (CDE) - Clinfowiki (n.d.) https://clinfowiki.org/wiki/index.php/Common_Data_Element_(CDE). Accessed 3 Apr 2018

da Veiga Leprevost F, Grüning BA, Aflitos SA, Röst HL, Uszkoreit J, Barsnes H, Vaudel M et al (2017) BioContainers: an open-source and community-driven framework for software standardization. Bioinformatics 33(16):2580–82

Danciu I, Cowan JD, Basford M, Wang X, Saip A, Osgood Susan, Shirey-Rice Jana, Kirby Jacqueline, Harris Paul A (2014) Secondary use of clinical data: the vanderbilt approach. J Biomed Inform 52(December):28–35

Drysdale R, Cook CE, Petryszak R, Baillie-Gerritsen V, Barlow M, Gasteiger E, Gruhl F et al (2020) The ELIXIR core data resources: fundamental infrastructure for the life sciences. Bioinformatics 36(8):2636–2642

Garza M, Fiol GD, Tenenbaum J, Walden A (2016) Evaluating common data models for use with a longitudinal community registry. J Biomed Inform 64(December):333–341

Goff SA, Vaughn M, McKay S, Lyons E, Stapleton AE, Gessler D et al (2011) The iPlant collaborative: cyberinfrastructure for plant biology. Front Plant Sci 2(July):34

Griffin PC, Khadake J, LeMay KS, Suzanna E, Lewis S, Orchard A, Pask B Pope, et al (2017) Best practice data life cycle approaches for the life sciences. F1000Research 6(August):1618

Grossman RL, Allison P, Heath V, Ferretti HE, Varmus DR, Lowy WA, Kibbe, Staudt LM (2016) Toward a shared vision for cancer genomic data. N Engl J Med 375(12):1109–1112

Gruening B, Sallou O, Moreno P, Felipe da Veiga Leprevost, Hervé Ménager, Dan Søndergaard, Hannes Röst, et al (2018) Recommendations for the packaging and containerizing of bioinformatics software. F1000Research 7(June). https://doi.org/10.12688/f1000research.15140.2

Harvey LA (2018) REDCap: web-based software for all types of data storage and collection. Spinal Cord. https://doi.org/10.1038/s41393-018-0169-9

Herzinger S, Gu W, Satagopam V, Eifes S, Rege K, Barbosa-Silva A, Schneider R, eTRIKS Consortium (2017) SmartR: an open-source platform for interactive visual analytics for translational research data. Bioinformatics 33(14):2229–2231

Hoogstrate Y, Zhang C, Senf A, Bijlard J, Hiltemann S, van Enckevort D, Susanna Repo, et al (2016) Integration of EGA Secure Data Access into Galaxy. F1000Research 5(December). https://doi.org/10.12688/f1000research.10221.1

Klann JG, Matthew AH, Joss KE, Murphy SN (2019) Data model harmonization for the all of us research program: transforming i2b2 data into the OMOP common data model. PLoS One 14(2): e0212463

Krishnaswamy S, Madhan Mohan T (2016) The largest distributed network of bioinformatics centres in the world: biotechnology information system network (DBT-BTISNET). Curr Sci. https://doi.org/10.18520/cs/v110/i4/556-561

Landis D, Courtney W, Dieringer C, Kelly R, King M, Miller B, Wang R, Wood D, Turner JA, Calhoun VD (2016) COINS data exchange: an open platform for compiling, curating, and disseminating neuroimaging data. NeuroImage 124(Pt B):1084–88

Malone J, Stevens R, Jupp S, Hancocks T, Parkinson H (2016) Ten simple rules for selecting a bio-ontology. PLoS Comput Biol 12(2):e1004743

Merchant N, Lyons E, Goff S, Vaughn M, Ware D, Micklos D (2016) The iPlant collaborative: cyberinfrastructure for enabling data to discovery for the life sciences. PLoS Biol 14(1):e1002342

Mohr C, Friedrich A, Wojnar D, Kenar E, Polatkan AC, Codrea MC, Czemmel S, Kohlbacher O, Nahnsen S (2018) qPortal: a platform for data-driven biomedical research. PLoS One 13(1):e0191603

Murphy S, Wilcox A (2014) Mission and sustainability of informatics for integrating biology and the bedside (i2b2). EGEMS (Washington DC) 2(2):1074

Musen MA, Carol A, Bean K-H, Cheung M, Dumontier KA, Durante Olivier Gevaert, Gonzalez-Beltran Alejandra et al (2015) The center for expanded data annotation and retrieval. J Am Med Inform Assoc: JAMIA 22(6):1148–52

Navale V, Bourne PE (2018) Cloud computing applications for biomedical science: a perspective. PLoS Comput Biol 14(6):e1006144

Navale V, Ji M, Vovk O, Misquitta L, Gebremichael T, Garcia A, Fann Y, and Matthew McAuliffe (2019) Development of an informatics system for accelerating biomedical research. F1000Research. https://doi.org/10.12688/f1000research.19161.1

Payakachat N, Tilford JM, Ungar WJ (2016) National Database for Autism Research (NDAR): big data opportunities for health services research and health technology assessment. PharmacoEconomics 34(2):127–138

Ravagli C, Pognan F, Marc P (2017) OntoBrowser: a collaborative tool for curation of ontologies by subject matter experts. Bioinformatics 33(1):148–149

Sansone S-A, McQuilton P, Rocca-Serra P, Gonzalez-Beltran A, Izzo M, Lister AL, Thurston Milo, FAIRsharing Community, (2019) FAIRsharing as a community approach to standards, repositories and policies. Nat Biotechnol 37(4):358–67

Scheufele E, Aronzon D, Coopersmith R, McDuffie MT, Kapoor M, Uhrich CA, Avitabile JE, Liu J, Housman D, Palchuk MB (2014) tranSMART: An open source knowledge management and high content data analytics platform. AMIA Joint Summits on Translational Science Proceedings. AMIA Joint Summits on Translational Science 2014 (April):96–101

Schneider M, Victoria PC, Griffin S, Tyagi M, Flannery S, Dayalan S, Gladman N, Watson-Haigh et al (2017) Establishing a distributed national research infrastructure providing bioinformatics support to life science researchers in Australia. Brief Bioinform 20(2):384–389

Sheehan J, Hirschfeld S, Foster E, Ghitza U, Goetz K, Karpinski J, Lang L et al (2016) Improving the value of clinical research through the use of common data elements. Clin Trails 13(6):671–676

Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ et al (2007) The OBO foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol 25(11): 1251–1255

Tauch A, Al-Dilaimi A (2019) Bioinformatics in Germany: toward a national-level infrastructure. Brief Bioinform 20(2):370–374

Toga AW (2017) The Global Alzheimer's Association Interactive Network (GAAIN). Alzheimers Dement. https://doi.org/10.1016/j.jalz.2017.07.025

Wilkinson MD, Dumontier M, Jsbrand Jan I, Aalbersberg G, Appleton M, Axton A, Baak N et al (2016) The FAIR guiding principles for scientific data management and stewardship. Sci Data 3(March): 160018

Zhang C, Bijlard J, Staiger C, Scollen S, van Enckevort D, Hoogstrate Y, Senf A et al (2017) Systematically linking tranSMART, galaxy and EGA for reusing human translational research data. F1000Research 6(August). https://doi.org/10.12688/f1000research.12168.1