Case report

BigMouth: development and maintenance of a successful dental data repository

Muhammad F. Walji1, Heiko Spallek2, Krishna Kumar Kookal1, Jane Barrow3, Britta Magnuson4, Tamanna Tiwari5, Udochukwu Oyoyo6, Michael Brandt7, Brian J. Howe8, Gary C. Anderson9, Joel M. White10, and Elsbeth Kalenderian3,10,11

1Department of Diagnostics and Biomedical Sciences. School of Dentistry, University of Texas Health Science Center at Houston, Houston, Texas, USA, 2Faculty of Dentistry. The University of Sydney, Sydney, Australia, 3Office of Global and Community Health. Harvard School of Dental Medicine, Boston, Massachusetts, USA, 4Department of Diagnostic Sciences. Tufts School of Dental Medicine, Boston, Massachusetts, USA, 5Department of Community Dentistry & Population Health. University of Colorado School of Dental Medicine, Aurora, Colorado, USA, 6Office of Dental Education Services. Loma Linda University School of Dentistry, Loma Linda, California, USA, 7Office of Information Resources. University of Buffalo School of Dental Medicine, Buffalo, New York, USA, 8Department of Family Dentistry. University of Iowa College of Dentistry and Dental Clinics, Iowa City, Iowa, USA, 9Department of Developmental and Surgical Sciences. University of Minnesota School of Dentistry, Minneapolis, Minnesota, USA, 10Department of Preventive and Restorative Dental Science. School of Dentistry, University of California at San Francisco, San Francisco, California, USA, and 11Department of Dental Management Sciences. School of Dentistry, University of Pretoria, Pretoria, South Africa

Muhammad F. Walji and Heiko Spallek contributed equally to this work.

Corresponding Author: Muhammad F. Walji, PhD, Department of Diagnostics and Biomedical Sciences, University of Texas Health Science Center at Houston, School of Dentistry, 7500 Cambridge St, SOD 4184, Houston, TX 77054, USA; Muhammad.f.walji@uth.tmc.edu

Received 16 July 2021; Revised 10 December 2021; Editorial Decision 28 December 2021; Accepted 20 January 2022

ABSTRACT

Few clinical datasets exist in dentistry to conduct secondary research. Hence, a novel dental data repository called BigMouth was developed, which has grown to include 11 academic institutions contributing Electronic Health Record data on over 4.5 million patients. The primary purpose for BigMouth is to serve as a high-quality resource for rapidly conducting oral health-related research. BigMouth allows for assessing the oral health status of a diverse US patient population; provides rationale and evidence for new oral health care delivery modes; and embraces the specific oral health research education mission. A data governance framework that encouraged data sharing while controlling contributed data was initially developed. This transformed over time into a mature framework, including a fee schedule for data requests and allowing access to researchers from noncontributing institutions. Adoption of BigMouth helps to foster new collaborations between clinical, epidemiological, statistical, and informatics experts and provides an additional venue for professional development.

Key words: dentistry, Research Patient Data Repositories, learning healthcare system, data governance
**INTRODUCTION**

Each year in the United States, over 195,000 dental practitioners provide care to more than 127 million patients.\(^1,2\) Despite significant advances over time, researchers have had limited access to oral health datasets. While different Research Patient Data Repositories (RPDRs) exist, they rarely contain information on oral health that is associated with chronic conditions.\(^3\) To answer critical oral health-related research questions, investigators often rely on small local datasets, which are difficult to generalize. Alternatively, data are sometimes obtained from third-party payers (e.g., dental insurance companies). The usefulness of these data may be limited as they are focused primarily on billed services, and many patients self-pay for dental care without involving a third party. The National Health and Nutritional Examination Survey\(^4\) and Behavioral Risk Factor Surveillance System\(^5\) are large oral health datasets, providing a view of the population’s dental status but contain limited information on dental diagnoses and actual treatments received. Linked datasets from medical and dental Electronic Health Records (EHRs) are also sorely lacking, impairing the ability to investigate relationships between oral health and general health.\(^6,7\) Recognizing this conundrum, we developed a centralized dental data repository using the i2b2 platform,\(^8\) called BigMouth.\(^7\) BigMouth was successfully launched in August 2012 with data on 1.1 million patients derived from dental EHRs of 4 dental schools—all members of the Consortium of Oral Health Research and Informatics (COHRI).\(^9,10\) Less than a decade later, BigMouth has grown into a formidable dental RPDR with 11 academic institutions contributing data on over 4.507 million patients (see Table 1) with diverse geographic coverage (see Figure 1).

**MAIN PURPOSE OF BIGMOUTH**

The objectives of BigMouth include informing the feasibility of research studies, executing informatic, population health, and observational studies, supporting quality improvement efforts, participation in data-driven research networks, and identifying clinical trial cohorts for recruitment. However, as a dental RPDR, BigMouth also fulfills several other purposes.

Assessment of oral health status quo

BigMouth’s data on 4.5 million patients distributed throughout the United States provide a remarkable window into the oral health status of a diverse patient population. BigMouth has been used to support, or refute, hypotheses of research proposals by generating preliminary data for funding applications and has therefore influenced the direction of oral health research. Moreover, BigMouth has been used to assess the quality of health intervention for specific patient populations, for example, providing dental care to women while pregnant,\(^11\) patients with diabetes,\(^12\) or children receiving sealants.\(^13,14\) Importantly, to date, the 11 participating BigMouth institutions are all academic—but vary in clinic size and resources for conducting research, and the results generated by BigMouth studies have started to illustrate that dental treatments in these educational settings are safe, effective, and cost-efficient.\(^15,16\)

Research education mission

BigMouth provides a secure environment in which oral health researchers at all levels (faculty, dental students, specialty trainees, master-level students, and doctoral-level students) have rapid access to a large dataset for analysis. Crucially, BigMouth serves as a “sandbox” to learn how to conduct clinical research using “real-world data,” appreciating all their limitations, including data availability, data quality, and challenges of electronic phenotyping.\(^17\)

Research priorities

By pooling datasets, BigMouth presents a more complete picture of types of patients. More specifically, BigMouth facilitates the study of oral manifestations of rare diseases that can lead to better understanding of the factors that affect more common diseases. Such rare diseases have a prevalence of fewer than 5 cases per 10,000 population\(^18\) with a 15% manifestation with oral-facial symptoms. As it takes on average 7 years to diagnose a rare disease, dentists can significantly influence identifying a rare disease by its oral symptomatology and help manage oral and overall quality of life.\(^19\) Infrequent diseases with oral health syndromes\(^20\) may be easier to locate in dental EHRs; however, individual institutions may not have enough patients with the disease for meaningful research. Another critical priority of BigMouth is the connection of oral health with general health. Information exchange between medical and dental EHRs is sorely limited because of ongoing interoperability issues, lack of consensus of what key components should be recorded in a patient’s record, and lack of documentation standards for dental EHRs.\(^21,22\) BigMouth includes medical history data and

**Table 1. BigMouth data elements by contributing institutions/site**

| Data Site       | Demographics | Diagnoses | Forms | Insurance | Odontogram | Periodontal Charts | Practice | Medications | Procedures |
|-----------------|--------------|-----------|-------|-----------|------------|--------------------|----------|-------------|------------|
| UT Houston      | 430 189      | 106 723   | 160 211 | 59 992    | 397 847    | 53 500             | 239 172  | 50 626      | 234 482    |
| UCSF            | 993 959      | 261 980   | 121 888 | 126 537   | 844 375    | 79 931             | 571 218  | 42 558      | 635 216    |
| HSDM            | 97 838       | 28 440    | 43 687 | 26 344    | 88 675     | 21 403             | 53 441   | 15 297      | 55 824     |
| Tufts           | 425 100      | 0         | 127 228 | 122 673   | 379 490    | 83 898             | 241 787  | 36 502      | 245 925    |
| U Pitt          | 291 648      | 0         | 103 591 | 71 319    | 256 299    | 38 531             | 115 668  | 36 328      | 117 202    |
| U Michigan      | 484 781      | 33 780    | 133 216 | 227 235   | 420 802    | 61 358             | 307 904  | 46 009      | 308 348    |
| UC Denver       | 167 180      | 14 714    | 48 986 | 44 644    | 142 394    | 34 117             | 99 143   | 41 096      | 105 688    |
| Loma Linda      | 482 526      | 22 961    | 195 839 | 115 082   | 393 961    | 67 686             | 207 089  | 32 584      | 210 171    |
| U Buffalo       | 231 398      | 22 741    | 21 317 | 9903      | 217 777    | 98 222             | 23 812   | 12 021      | 25 065     |
| U Iowa          | 801 739      | 989       | 464 412 | 325 697   | 665 953    | 53 844             | 518 717  | 142 689     | 512 259    |
| U Minnesota     | 101 274      | 3689      | 101 274 | 77 535    | 101 274    | 46 393             | 95 410   | 49 836      | 97 019     |
| Total patients  | 4 507 632    | 495 926   | 1 478 902 | 1 207 161 | 3 909 247 | 542 808             | 2 473 361 | 525 546     | 2 547 199 |
data on medications, as such allowing for investigation of the oral health-overall health relationship. Importantly, BigMouth will also allow for the building of a sustainable dental learning health system focused on providing patients with safe and effective oral health care.

GOVERNANCE

As part of the formation of BigMouth, a data governance framework was developed that encouraged data sharing while allowing control of contributed data. BigMouth was originally conceived and remains as a single central repository, where all sites deposit their data. While there have been discussions about moving to a federated model, the complexity of requiring contributing sites to host their own i2b2 instances was a major barrier. Over time, changes were made to encourage site participation, access of data, execution of research studies, and sustainability. Specifically, data from beyond the original dental EHR (axiUm, Vancouver, Canada) were accepted by BigMouth, allowing for more sites to participate. This is important as several institutions are moving to Epic (Epic, Verona, WI). The Governance Committee permits noncontributing sites to receive data from BigMouth as long as those queries are for noncommercial purposes. The decision to limit access to data in BigMouth to nonprofits was based on advice from institutional legal representatives to ensure there was no perception that patient data were being monetized. As our institutions are gaining experience and developing formal policies for collaborating with for-profit entities for discovery using EHR data, we expect to revisit our restrictions on limiting access.

A 3-year National Library of Medicine resource development grant (G08LM010075) originally supported the formation of BigMouth and sharing of data from the 4 founding institutions. Supporting sustainability, a one-time setup fee of $10 000 has been implemented for new sites, and all sites pay an annual fee of $2500 which supports updating the repository on a quarterly basis. Although there is no cost for participating sites to query summary data using i2b2 web interface, there is now a fee schedule, based on complexity, for requests that require extracting data with costs for non-contributing members being higher than for contributors.

OPERATIONS

Conducting research using BigMouth: approaches and lessons learned

Researchers aspiring to use BigMouth data submit a proposal using a predefined template. Each contributing site has a representative who will first independently review the proposal based on scientific merit, potential overlap with other approved BigMouth projects, and if the institution agrees to share data for the proposed research. To date, the BigMouth committee has formally reviewed and approved 18 research proposals. In order to minimize the back and forth, researchers are now invited to join the committee discussion which has led to a robust process with faster approvals.
Our vision is to incorporate dental imaging data (2d and 3d) and mandate the use of standardized diagnostic terminologies by all contributing institutions. Connecting to mobile health apps that collect patient-reported outcomes as well as patient-reported experience measures is also crucial aspiring toward a more holistic definition of oral health.

3. Using BigMouth to improve oral health care: Dentistry lags behind the medical profession in fostering the uptake of research-informed treatments. The use of computerized knowledge management, for example, in the form of audit and feedback and clinical decision support has been introduced in the oral health arena and will undoubtedly facilitate alignment of every day dental practice with evidence-based guidelines. Hence, BigMouth is positioned to play an important role in getting dental teams and clinics to adopt and consistently use evidence-based oral health guidelines and will be a catalyst for the transition from payment-focused care to culturally sensitive, effective, and high-value oral health care.

We have many challenges ahead, including the expansion of BigMouth to other contributing institutions. While we have developed a process for rapidly incorporating data from the axiUm, EHR, we
will need to develop scalable approaches for incorporating data from other platforms. We also strive to include contributing institutions globally, which will require an understanding of legal and policy issues of sharing patient data across borders. While dental institutions have been willing to share patient data, we expect more challenges for connecting or incorporating data from the patient’s medical record. Returning to the FDI definition, we aspire to embrace this holistic definition and finally help move dentistry from treating disease to treating a person with disease.

CONCLUSION
Visionary leadership, combined with a strong governance approach to data sharing, has made the large-scale dental data repository, BigMouth, a reality. Initial federal funding and ongoing efforts to develop sustainability have supported researchers’ efforts to mine data otherwise not available to advance dental research.

FUNDING
This work was supported initially by the National Library of Medicine (grant number G08LM010075).

AUTHOR CONTRIBUTIONS
MFW, EK, and HS conceptualized the manuscript and wrote an initial draft. All authors expanded, proofread, and substantially edited the manuscript. MFW, EK, and JMW secured funding for the project.

ACKNOWLEDGMENTS
We thank Dean John Valenza at UTHealth School of Dentistry for his ongoing support of and belief in BigMouth. We also thank Joseph Applegate, Susan Guerero, and Dr Elmer Bernstam from the UTHealth School of Biomedical Informatics for their continued contributions to maintaining BigMouth. We also thank Roger Gillie from the University of Michigan School of Dentistry and Dr Charles Sfeir from the University of Pittsburgh School of Dental Medicine for representing their institutions.

CONFLICT OF INTEREST STATEMENT
None declared.

DATA AVAILABILITY
The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES
1. Munson B, Vujic M. Supply of Dentists in the United States Is Likely to Grow. Health Policy Institute Research Brief. Chicago, Ill.: American Dental Association; 2014.
2. Franklin A, Gantela S, Shifarraw S, et al. Dashboard visualizations: supporting real-time throughput decision-making. *J Biomed Inform* 2017; 71: 211–21.
3. Institute of Medicine and National Research Council. *Improving Access to Oral Health Care for Vulnerable and Underserved Populations*. Washington, DC: The National Academies Press; 2011: 296.
4. Zipf G, Chiappa M, Porter K, Ostchega Y, Lewis B, Dostal J. National health and nutrition examination survey: plan and operations, 1999-2010. *Vital Health Stat Ser 1 Programs Collect Proced* 2013; (56): 1–37.
5. Remington PL, Smith MY, Williamson DF, Anda RF, Gentry EM, Hoge lin GC. Design, characteristics, and usefulness of state-based behavioral risk factor surveillance: 1981–87. *Public Health Rep* 1988; 103 (4): 366.
6. Boland MR, Hripcsak G, Albers DJ, et al. Discovering medical conditions associated with periodontitis using linked electronic health records. *J Clin Periodontol* 2013; 40 (5): 474–82.
7. Kalenderian E, Halaman J, Spallek H. An EHR with teeth. *Appl Clin Inform* 2016; 7 (2): 425–9.
8. Murphy SN, Weber G, Mendis M, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc* 2010; 17 (2): 124–30.
9. Waliﬁ JS, Kalenderian E, Stark PC, et al. BigMouth: a multi-institutional dental data repository. *J Am Med Inform Assoc: JAMIA* 2014; 21 (6): 1136–40.
10. Stark PC, Kalenderian E, White JM, et al.; Consortium for Oral Health-Related Informatics (COHRI). Consortium for oral health-related informatics: improving dental research, education, and treatment. *J Dent Educ* 2010; 74 (10): 1051–65.
11. Neumann A, Obadan-Udoh E, Bangar S, et al. Number of pregnant women at four dental clinics and the care they received: a dental quality eMeasure evaluation. *J Dent Educ* 2019; 83 (10): 1158–65.
12. Neumann A, Kalenderian E, Ramoni R, et al. Evaluating quality of dental care among patients with diabetes: adaptation and testing of a dental quality measure in electronic health records. *J Am Dent Assoc* 2017; 148 (9): 634–43.e1.
13. Kumar SV, Bangar S, Neumann A, et al. Assessing the validity of existing dental sealant quality measures. *J Am Dent Assoc* 2018; 149 (9): 756–64.e1.
14. Yansane A, Lisl S, Dawda D, et al. Increasing value, reducing waste: tailoring the application of dental sealants according to individual caries risk. *J Public Health Dent* 2020; 80 (Supp 2): 58–16.
15. Tokede O, Ramoni RB, Patton M, Da Silva JD, Kalenderian E. Clinical documentation of dental care in an era of electronic health record use. *J Evid Based Dent Pract* 2016; 16 (3): 154–60.
16. Kalenderian E, Obadan-Udoh E, Yansane A, et al. Feasibility of electronic health record-based triggers in detecting dental adverse events. *Appl Clin Inform* 2018; 9 (3): 646–53.
17. Rogers JR, Lee J, Zhou Z, Cheung YK, Hripcsak G, Weng C. Contemporary use of real-world data for clinical trial conduct in the United States: a scoping review. *J Am Med Inform Assoc: JAMIA* 2021; 28 (1): 144–54.
18. Molina-Garcia A, Castellanos-Cosano L, Machuca-Portillo G, Posada-de la Paz M. Impact of rare diseases in oral health. *Med Oral Patol Oral Cir Bucal* 2016; 21 (5): e587–94.
19. Wiemann S, Frenzel Baudisch N, Jordan RA, Kleinheinz J, Hanisch M. Oral symptoms and oral health-related quality of life in people with rare diseases in Germany: a cross-sectional study. *Int J Environ Res Public Health* 2018; 15 (7): 1493.
20. Parker SE, Mait CT, Canfield MA, et al. Updated national birth prevalence estimates for selected birth defects in the United States, 2004-2006. *Birth Defects Res A Clin Mol Teratol* 2010; 88 (12): 1008–16.
21. Simon I, Obadan-Udoh E, Yansane A, et al. Improving oral-systemic healthcare through the interoperability of electronic medical and dental records: an exploratory study. *Appl Clin Inform* 2019; 10 (3): 367–76.
22. Schleyer T, Spallek H, Hernandez P. A qualitative investigation of the content of dental paper-based and computer-based patient record formats. *J Am Med Inform Assoc: JAMIA* 2007; 14 (4): 515–26.
23. Messing M, Souza LC, Cavalla F, et al. Investigating potential correlations between endodontic pathology and cardiovascular diseases using epidemiological and genetic approaches. *J Endod* 2019; 45 (2): 104–10.
24. Neumann A, Kumar S, Bangar S, et al. Tobacco screening and cessation efforts by dental providers: a quality measure evaluation. *J Public Health Dent* 2019; 79 (2): 93–101.
25. Adams BJ. Establishing personal identification based on specific patterns of missing, filled, and unrestored teeth. J Forensic Sci 2003; 48 (3): 487–96.

26. Sajadi A. Clinical Accuracy of Pre-doctoral Periodontal Education in the US: An Evaluation of the Generalist and Specialist Teaching Models on Periodontal Education. Houston, TX: The University of Texas School of Dentistry at Houston; 2015.

27. Tran DT, Gay JG, Du XL, et al. Partial-mouth periodontal examination protocol for estimating periodontitis extent and severity in a US population. Clin Exp Dent Res 2016; 2 (1): 73–9.

28. Walji M, Ramoni R, Simmons K, et al. Using EHRs and big data for measuring oral healthcare quality. J Dent Res 2017; 96 (A): 0553.

29. Kumar S, Obadan-Udoh E, Yansane A, et al. Monitoring sealant treatment practices using EHR-based quality measures. J Dent Res 2018; 97 (A): 2849700.

30. Kalenderian E, Obadan-Udoh E, Neumann A, et al. Mining the electronic health record for dental adverse events. J Dent Res 2018; 97 (A): 2853847.

31. Mc Cord KA, Hemkens LG. Using electronic health records for clinical trials: where do we stand and where can we go? CMAJ 2019; 191 (5): E128–33.

32. Adibi S, Li M, Salazar N, et al. Medical and dental electronic health record reporting discrepancies in integrated patient care. JDR Clin Trans Res 2020; 5 (3): 278–83.

33. Centola D. How Behavior Spreads: The Science of Complex Contagions. Princeton, NJ: Princeton University Press; 2018.

34. Federation FWD. FDI's Definition of Oral Health. Geneva, Switzerland: FDI. https://www.fdiworlddental.org/fdis-definition-oral-health. Accessed January 1, 2022.

35. Hansen GM, Egeberg A, Holmstrup P, Hansen PR. Relation of periodontitis to risk of cardiovascular and all-cause mortality (from a Danish nationwide cohort study). Am J Cardiol 2016; 118 (4): 489–93.

36. Jeffcoat MK, Jeffcoat RL, Gladowski PA, Bramson JB, Blum JJ. Impact of periodontal therapy on general health: evidence from insurance data for five systemic conditions. Am J Prev Med 2014; 47 (2): 166–74.

37. Wagle NW. Implementing patient-reported outcome measures. NEJM Catalyst. http://catalyst.nejm.org/implementing-proms-patient-reported-outcome-measures. Accessed January 1, 2022.

38. Wittneben JG, Wismeijer D, Bragger U, Joda T, Abou-Ayash S. Patient-reported outcome measures focusing on aesthetics of implant- and tooth-supported fixed dental prostheses: a systematic review and meta-analysis. Clin Oral Implants Res 2018; 29 (Suppl 16): 224–40.

39. Yao CJ, Cao C, Bornstein MM, Matthes N. Patient-reported outcome measures of edentulous patients restored with implant-supported removable and fixed prostheses: a systematic review. Clin Oral Implants Res 2018; 29 (Suppl 16): 241–54.

40. Hodson M, Andrew S, Roberts CM. Towards an understanding of PREMs and PROMS in COPD. Breathe 2013; 9 (5): 358–64.

41. O’Donnell JA, Modesto A, Oakley M, Polk DE, Valappil B, Spallek H. Sealants and dental caries: insight into dentists’ behaviors regarding implementation of clinical practice recommendations. J Am Dent Assoc 2013; 144 (4): e24–30.

42. Rindal DB, Rush WA, Boyle RG. Clinical inertia in dentistry: a review of the phenomenon. J Contemp Dent Pract 2008; 9 (1): 113–21.

43. van der Sanden WJ, Nienhuijs ME, Mettes TG. The role of guidelines and systematic reviews in oral healthcare. Ned Tijdschr Tandheelkd 2007; 114 (4): 179–86.

44. Loffler C, Bohmer F. The effect of interventions aiming to optimise the prescription of antibiotics in dental care – a systematic review. PloS One 2017; 12 (11): e0188061.

45. Elouafkaoui P, Young L, Newlands R, et al. An audit and feedback intervention for reducing antibiotic prescribing in general dental practice: the RAPID cluster randomised controlled trial. PLOS Med 2016; 13 (8): e1002115.

46. Tianwiwat S, Hintao J, Chongsuvivatwong V, Thitasomakul S. Improvement of dental nurses awareness of school dental sealant quality following the audit and feedback system: first phase of implementation. Editorium J Dent 2015; 2: 7–14.

47. Titler MG, Titler P. The evidence for evidence-based practice implementation. In: Hughes RG, ed. Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville, MD: Agency for Healthcare Research and Quality, 2008: 1-113–1-132.

48. Vonnea-Griffin A, Fellows JL, Rindal DB, Barasch A, Gilbert GH, Safford MM. Pay for performance: will dentistry follow? BMC Oral Health 2010; 10: 9.