

UC Davis

UC Davis Previously Published Works

Title

Implementing partnership-driven clinical federated electronic health record data sharing networks

Permalink

<https://escholarship.org/uc/item/25k732mj>

Authors

Stephens, Kari A
Anderson, Nicholas
Lin, Ching-Ping
et al.

Publication Date

2016-09-01

DOI

10.1016/j.ijmedinf.2016.05.008

Peer reviewed



Implementing partnership-driven clinical federated electronic health record data sharing networks



Kari A. Stephens^{a,b,c,*}, Nicholas Anderson^d, Ching-Ping Lin^e, Hossein Estiri^c

^a Department of Psychiatry & Behavioral Sciences, University of Washington, Box 356560, Seattle, WA 98195, United States

^b Department of Biomedical Informatics & Medical Education, University of Washington, Box 358051, Seattle, WA 98109, United States

^c Institute of Translational Health Sciences, University of Washington, Box 358051, Seattle, WA 98109, United States

^d Department of Pathology and Laboratory Medicine, University of California Davis, Davis, CA 95616, United States

^e Global REACH, Medical School, University of Michigan, 5113 Medical Science Building I, 1301 Catherine St, Ann Arbor, MI 48109-5611, United States

ARTICLE INFO

Article history:

Received 11 August 2015

Received in revised form 4 January 2016

Accepted 30 May 2016

Keywords:

Information systems

Data sharing

Federated networks

Implementation

Electronic health records

ABSTRACT

Objective: Building federated data sharing architectures requires supporting a range of data owners, effective and validated semantic alignment between data resources, and consistent focus on end-users. Establishing these resources requires development methodologies that support internal validation of data extraction and translation processes, sustaining meaningful partnerships, and delivering clear and measurable system utility. We describe findings from two federated data sharing case examples that detail critical factors, shared outcomes, and production environment results.

Methods: Two federated data sharing pilot architectures developed to support network-based research associated with the University of Washington's Institute of Translational Health Sciences provided the basis for the findings. A spiral model for implementation and evaluation was used to structure iterations of development and support knowledge share between the two network development teams, which cross collaborated to support and manage common stages.

Results: We found that using a spiral model of software development and multiple cycles of iteration was effective in achieving early network design goals. Both networks required time and resource intensive efforts to establish a trusted environment to create the data sharing architectures. Both networks were challenged by the need for adaptive use cases to define and test utility.

Conclusion: An iterative cyclical model of development provided a process for developing trust with data partners and refining the design, and supported measureable success in the development of new federated data sharing architectures.

© 2016 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The broad adoption of electronic health record systems (EHRs) and efforts to align data across disparate EHRs have led to advancements in research to improve public health. But barriers to establish effective data sharing systems range across technical, motivational, economic, legal, political, and ethical issues [1]. Data sharing has an integral role in reducing the lag between research and clinical knowledge, products, and procedures that can improve human health [2]. Bi-directional data sharing between clinical care and research environments is crucial to advance improvements in patient care and overall population health and essential to a Learn-

ing Healthcare System [3]. But creating data sharing systems is complex and difficult.

Technical and methodological frameworks and guidelines for providing and integrating data sharing infrastructures across multiple distinct and disparate clinical environments can advance the ability for translational and comparative effectiveness research, and lead to meaningful use and sharing of medical data [4]. However, there are no systematic efforts to develop processes for creating data sharing architectures in public health environments [1]. Published accounts addressing builds of data sharing infrastructures lack any systematic application of well-established software development models. At present, implementation of data sharing systems are often supported by grant funding and require the development of broad engagement strategies between disparate environments. Sustainability of these systems often becomes a challenge after initial investments support creation [5]. Software

* Corresponding author at: Psychiatry & Behavioral Sciences, University of Washington, Box 356560 Seattle, WA 98195, United States.

E-mail address: kstephen@uw.edu (K.A. Stephens).

model applications to architecture builds may lead to better sustainability.

1.1. Background and significance

Previous efforts in developing methods and tools to support clinical data sharing for research lack access to high quality data sources [6–8]. Centralized approaches to data sharing are limited by the scope of the data that network partners typically authorize for sharing and the difficulty with keeping these data up to date [4]. Historically, limitations have also included uneven common terminology expertise, challenges of trust and feasibility, and concerns for privacy and security [4,9–17]. Storing data locally at partner sites and using federated approaches to support data sharing is attractive because they simplify privacy and security issues and clarify trust relationships [18]. However, no standard use of terminologies and other semantic alignment issues remain a challenge, regardless of a centralized versus federated model [19]. To date, large scale federated data sharing networks remain relatively scarce, though successes have been increasing in domain-specific networks such as Regional Health Information Organizations and cohort discovery pilots [19–22]. Growing concerns of enhanced HIPAA privacy laws may further limit data sharing efforts [23].

The expanding use of health information technology, driven through efforts such as the 2009 HITECH act and the meaningful use requirement of health information exchanges, has created the need for effective data sharing methods across organizations to target evaluation and implementation of evidence based, patient-centered clinical practices [24–26]. Methodological approaches to developing federated data sharing networks need to be testable and generalizable to multiple domains, users, and stakeholders. The NCATS Clinical Translational Science Award (CTSA) consortium has provided a fertile environment for building federated data sharing networks across a range of heterogeneous institutional and community based clinical environments with a focus on translational science.

1.2. Objective

We partnered across two network teams to implement and evaluate a software development model for building federated electronic health record clinical data sharing architectures. We describe the use of a common spiral model and the experience of developing two distinct architectures. Implementation of the spiral model centrally incorporated partnership building across different clinical data environments and addressed the crucial role of partnerships and disparate electronic medical record platforms and workflows.

2. Methods

2.1. Network development pilot projects

The common goal of our network pilot projects was to implement architectures for federated networks that could support research queries through a common set of terminologies and business processes. The Data QUery, Extraction, Standardization, Translation (Data QUEST) project focused on data sharing across primary care based electronic health record (EHR) data domains (i.e., demographics, visits, problem lists, medications, labs, diagnoses, tests, various medical metrics and findings, etc.) across six primary care organizations in Washington and Idaho [27]. Data QUEST is aimed to provide tools for sharing both de-identified and identified data in aggregate form and at the patient level. The Cross-Institutional Clinical Translational Research (CICTR) project

targeted sharing five broad data domains (i.e., demographics, medications, labs, diagnoses, and disposition data), with a common domain of diabetes across acute care settings at three academic institutions (University of Washington, University of California, San Francisco, and University of California, Davis) with a focus of sharing de-identified aggregated data [28]. Both projects used HIPAA guidance to define privacy handling of data prior to allowing research querying. Both projects supported approaches that describe and document the data provenance.

2.2. Procedure

Three primary categories of software models have been identified (free/open source software (FOSS), plan-driven, and agile) with little progress made at creating comprehensive reconciliation across these models [29]. However, recommendations for selecting an appropriate model include achieving a balance between agility and discipline [30]. The strength of FOSS lies in allowing stakeholders to address and refine a system based on individual priorities and resources. This model did not provide a feasible approach, given our partner sites must share resources and technical solutions to remain scalable in a diverse health data sharing architecture environment. Plan-driven or waterfall models lack iterative processes for achieving stakeholder engagement across cycles of development that provide flexibility, buy in, and adaptability. Agile methods are iterative but rely on quick “sprints” through the phases of development to produce working systems for evaluation, which require intensive development resources and evaluation resources (clinical and technical) from our partners that they did not have.

To balance agility and discipline, we chose Boehm’s spiral model, used across many commercial and defense projects, which included a focus on using a cyclic approach to grow a system’s degree of definition and implementation while laying out anchor point milestones to ensure stakeholder commitment to the defined solutions [31,32]. The spiral model was used to provide clear process to guide our architecture development and included cycles for iteration, incremental development, and the right level of risk management and cultural compatibility for our environment. We analyzed project activities, milestones, stakeholder priorities, and project documents using themes from Boehm’s spiral model of development, which included four main phases in the software development lifecycle, to define additional emerging themes. Each team, in partnership with project stakeholders, then reviewed and iterated on the emerging themes and charted the history of the project across the theme areas to develop initial project specific content for a draft spiral model. The resulting model was adopted within the Data QUEST and CICTR project teams, guiding biomedical informatics work within the projects. The model provided a frame to report and assess both individual project and cross project successes and challenges.

3. Results

3.1. The partnership-driven clinical federated (PCF) model

3.1.1. PCF model description

A generic spiral model for partnership-driven clinical federated (PCF) data sharing, based on Boehm’s spiral model for software development [31,32], emerged from our iterative and qualitative based methods (Fig. 1). This model identified four themes to anchor the iterative process of development: (1) developing partnerships, (2) defining system requirements, (3) determining technical architecture, and (4) conducting effective promotion and evaluation.

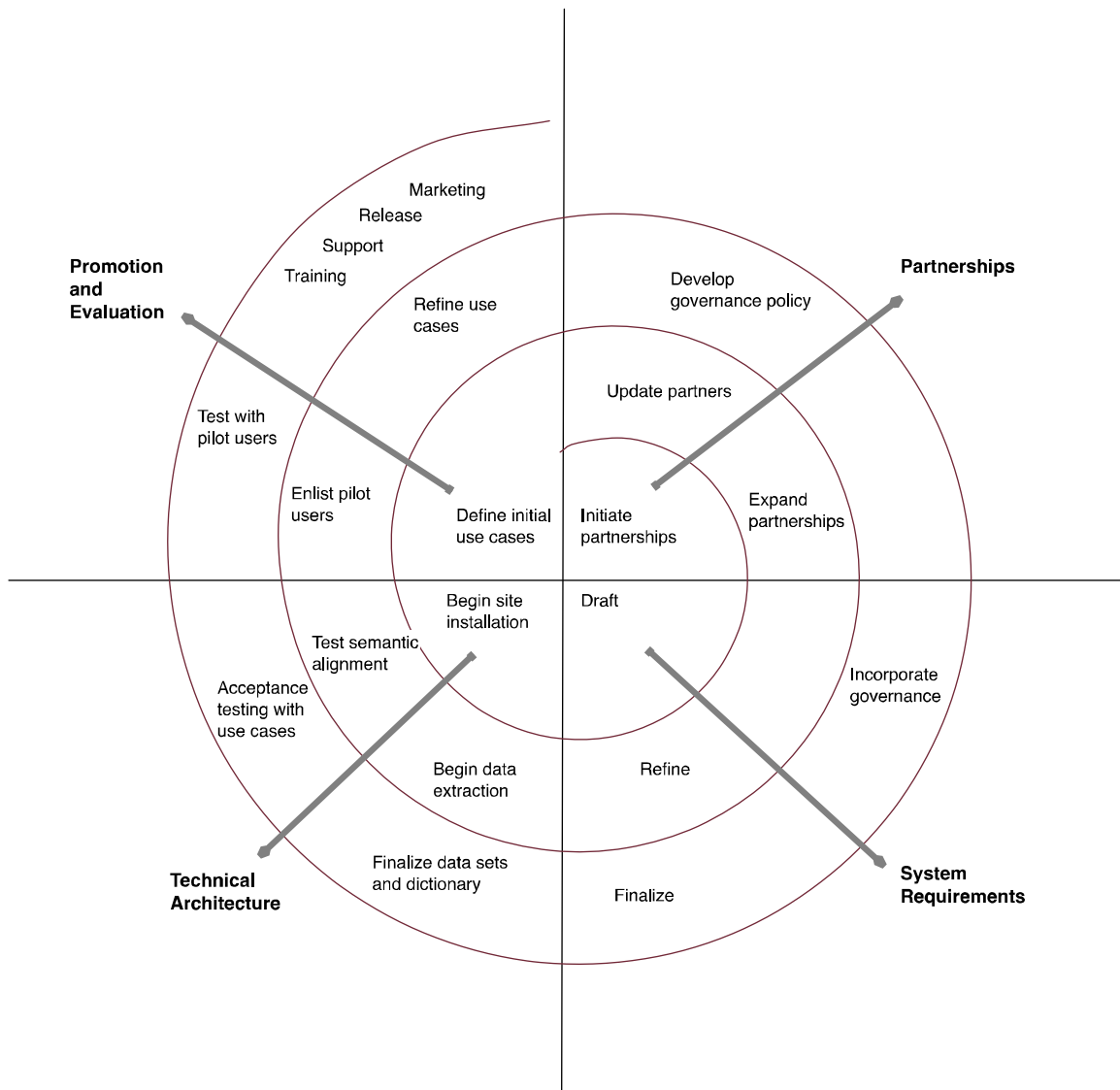


Fig. 1. The Partnership-Driven Clinical Federated (PCF) data sharing model illustrates four quadrants of themes used to define each iteration cycle of development.

The starting point was anchored in developing functional partnerships (Partnerships) that define boundaries and drive system definition. As the data sharing system was defined (System Requirements), implementation could occur (Technical Architecture), and finally impact was assessed (Promotion and Evaluation) to ensure the utility and impact of the data sharing network.

3.1.2. Cross-project model validity

As each theme progressed through a single iteration of the model, they informed the subsequent iterations and matured. Maturity of the model within each network occurred through iterations as well as across themes. Partnerships began internally with core project teams and progressed to recruitment and expansion to additional community partners, with final governance being addressed as maturity was reached. System requirements were initially drafted and subsequently refined as partnerships and technical architectures were developed. Use cases were initiated with the development of initial partnerships and evolved over time into meaningful use cases that addressed overall utility of the architecture. In tandem, pilot users were identified and training, support, release, and marketing efforts were developed and implemented to ensure system utility and sustainability. In general, the model

provided a useful framework for teams across projects to collaborate, report progress to each other, and share and iterate lessons learned.

3.2. Data query, extraction, standardization, translation (Data QUEST) project

3.2.1. Individual PCF model cycle iterations

The Data QUEST project conformed to the PCF model cycles (see Fig. 2 for a project specific model), iterating through four cycles. Anchor point milestones were developed and adjusted as needed, based on technical discoveries and partner requirements. The first cycle began with the team initiating partnerships within our CTSA partners, then moved into drafts of technical requirements, system testing of initially pre-defined technical architecture solutions, which subsequently failed, and initial definitions of use cases. We formed internal partnerships between all CTSA partners and convened regular meetings to discuss and evaluate adoption of the HMORN Virtual Data Warehouse (VDW) model [33], while the community engagement and biomedical informatics subgroups began developing feasibility study methods to determine selection of community based partners for the pilot. Development of feasibility

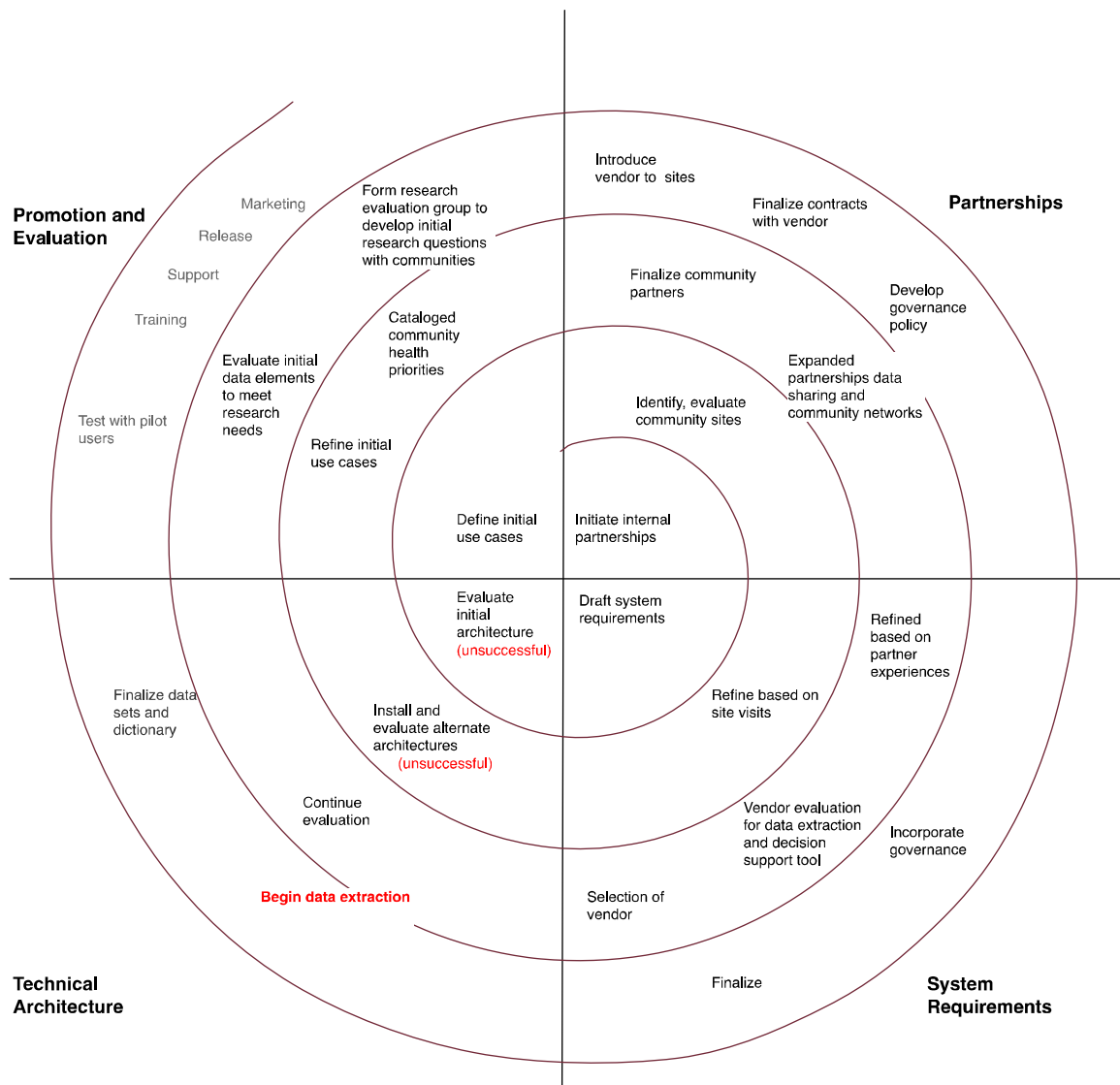


Fig. 2. PCF model applied to data QUEST detailing four cycles of iteration to mature the initial launch of the data sharing architecture. Technical architecture failures are highlighted in red.(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

methods resulted in the rejection of the VDW as a technical solution, due to the technical requirement for programming expertise among the community practice partners needed to work with the SAS based architecture. Our community practice partners reported that they did not have resources to support this level of on-site programming.

In the second cycle, we identified and approached initial community partnerships using the feasibility study methods (i.e., a semi-structured interview involving research readiness and technical capacity assessment), allowing for community partner input on technical requirements. A search was conducted to identify a replacement solution for the VDW, which led to e-PCRN [34,35]. The biomedical informatics team downloaded the e-PCRN software for pilot test and the test was unsuccessful, with the software deemed non-functional. Furthermore, the governance requirements dictated from the feasibility studies required on-site validation of all queries, which e-PCRN did not support and co-development of the software to add this feature was untenable. Use cases across the CTSA team and the community partners were iterated, cataloging community priorities for research and achieving buy in from partners for the utility and need for use case definition. This also led to redefining technical requirements to scale back functionality by

excluding a tool requirement to conduct self service queries against the federated data system.

In the third cycle, we solidified community partnerships and finalized selection of pilot partners for installation of the technical architecture and updated partners with the changes to the technical requirements to maintain continued buy in. National partnerships, most notably with the DARTNet Institute, were engaged and promoted selection of a for-profit vendor solution for extraction/translation/loading (ETL) tasks. Several vendors were evaluated and a final vendor was selected, based on their extensive expertise with multiple primary care based EHR vendor systems, ability to offer clinical decision support tools, success with working in Practice Based Research Network settings with small rural based primary care practices, and their existing relationship and good performance history with the DARTNet Institute [27]. Evaluation groups were formed to refine pilot use cases based on research priorities from the community partners.

In the fourth cycle, we introduced partners to the vendor and iterated and established contracts and governance (i.e., Memorandums of Understanding, Data Use Agreements, Business Associate Agreements, purchasing contracts) to allow initial installation of the technical architecture. System requirements were adjusted to

include vendor and partner requirements, and dictionary architectures were begun, with dictionary efforts designed to support marketing and dissemination of data sharing across the new network. After completion of the fourth cycle, the technical architecture was deemed operational and initial use case based extractions were performed.

Costs and resources used for the initial pilot build of Data QUEST included: (1) biomedical informatics personnel (i.e., faculty lead and dedicated research assistant), (2) community outreach CTSA faculty and staff (i.e., faculty lead and program manager), (3) CTSA subsidized staff resources (i.e., system architects and analysts for consultation), and (4) infrastructure contracts paid out to our vendor to conduct contracting and programming to establish servers and the nightly ETL process. Our partner sites also subsidized: (1) staff time to participate in working meets and establish permissions with our vendor to establish a server within their firewalls and (2) infrastructure support to house the server. The Data QUEST pilot project was designed and implemented within the CTSA 5 year cycle.

In the current state and subsequent CTSA cycle, the Data QUEST architecture is fully functional with daily refreshes of a federated set of clinical data repositories (CDR) across pilot sites, housed physically behind firewalls within the partner practices in SQL Server environments. In addition, data within the CDRs are stored in formats that semantically align with national partner networks within the DARTNet Institute, allowing for cross collaborations. The vendor and the DARTNet Institute have physical access to the CDRs and provide manual data extractions as needed, governed by agreements designed and executed within the development cycles. Several research projects to date, including local university and national collaboration based projects have been successfully conducted using the Data QUEST architecture. All data extractions are conducted using Business Associate Agreements and governance that provides honest brokerage, including compliance with Health Insurance Portability and Accountability Act (HIPAA) regulations, Data Use Agreements, and Data Transfer Agreements. Data remain owned by the local partners and data owners approve each extraction as they are requested, with the ability to opt out at any time. A tool (FindIT—Federated Information Dictionary Tool) [36] to catalogue data depth and breadth is under development targeting data visualization of the Data QUEST network to the research community, with an aim towards bridging researchers and community based practices and increasing use of the EHR data to facilitate translational research among the Data QUEST network partners. We have developed the architecture to support a centralized de-identified warehouse adopting the Observation Medical Outcomes Partnership (OMOP) ontology, harmonizing with other national data sharing architectures and plans to continue partner expansion of Data QUEST.

3.2.2. Cross-cycle observations

The Data QUEST project initiatives and priorities were accounted for across the four themes. Two technical architecture failures occurred, and rather than having progress stifled, the iterative nature of the model strengthened partnerships and supported refinement of feasible system requirements. The PCF model provided natural feedback loops between themes to coordinate and problem solve issues across themes. Initial partnership input was critical to the development of use cases and development of the final technical architecture. The team engaged in micro iterations when barriers occurred in testing the technical architecture solution, resulting in higher utilization of time resources needed to engage and mature partnerships.

The team identified several barriers requiring mediation within themes that required further micro iterations within cycles resulting in time delays, but ultimately did not compromise project

success. Limited partner expertise or knowledge in informatics, for both institutional and community partners and limited experience in being part of a multi-disciplinary institutional and community team were barriers to initially developing effective use cases. Partners struggled with limited resources due to overly burdened clinical environments and had limited experience with research practices. A primary outcome of the application of the PCF model was the critical role of early engagement in partnerships to mediate identified barriers through iteration. Despite real-world challenges to aligning partners and resources in a collaborative effort, iterations continued across cycles, allowing individual themes to mature.

3.3. Cross-Institutional clinical translational research (CICTR) project

3.3.1. Individual PCF model cycle iterations

The CICTR project conformed to the PCF model cycles (see Fig. 3 for a project specific model), iterating through three cycles with predefined anchor point milestones to move the project towards the first version release within a 20 month timeline dictated by the pilot funding requirements. The first cycle focused on defining stakeholder roles at the three partner sites, deploying common Informatics for Integrating Biology & the Bedside (i2b2) [37] based technical architectures and gaining appropriate regulatory approval. At the time, each of these sites had existing familiarity with i2b2, but not in developing environments that could provide appropriate hosting for large scale extracts from the local EHR systems. This first coordinated effort resulted in each of the sites establishing the i2b2 environments as common virtualized environments located within local server resources, and allowed for cross-site network and server configuration. Each site then implemented their local i2b2 server stack against an industry grade database system, either IBM DB2, Microsoft SQL Server, or Oracle 11. Each of these database systems were unique to local sites, and due to licensing restrictions and costs, requirements were defined that maintained these systems as local resources.

The second cycle built upon on the common deployed i2b2 architecture and partners iterated the access, common definition, and testing development for the four target data sources of demographics, diagnoses, medications and laboratory tests. In parallel, the SHRINE [38] network interfaces were installed and tests were defined to measure pilot ‘connectivity’ across the network. Use cases were collectively developed to test both within system and across system validity for the first two data sources. As the systems were being launched, each site devoted considerable effort to recruiting new domain stakeholders in the testing and demonstration of the functionality. An early challenge was in determining how to define and refine the use cases as to allow for review and measurement across sites by the stakeholders who were local to an individual site, in a way that would inform better processes for designing and testing the remaining data sources. By the end of the second phase, the network had completed basic validation of the first two data sources with common tests at each site while maintaining and growing partner engagement.

The third and final planned cycle sought to complete mapping of medication and laboratory data across the three sites, and through this requirement identified significant challenges for defining scope of the much larger mapping efforts involved in these data sources. As there were limited resources at each site, semi-automated tools were tested, which revealed challenges in cross-site evaluation of resulting mappings. A key project contribution was a tool developed at University of California, San Francisco for terminology mapping, which provided programmatic access to a standard medication coding system (RxNorm API). Through this, each site sought to map a common reference medication formu-

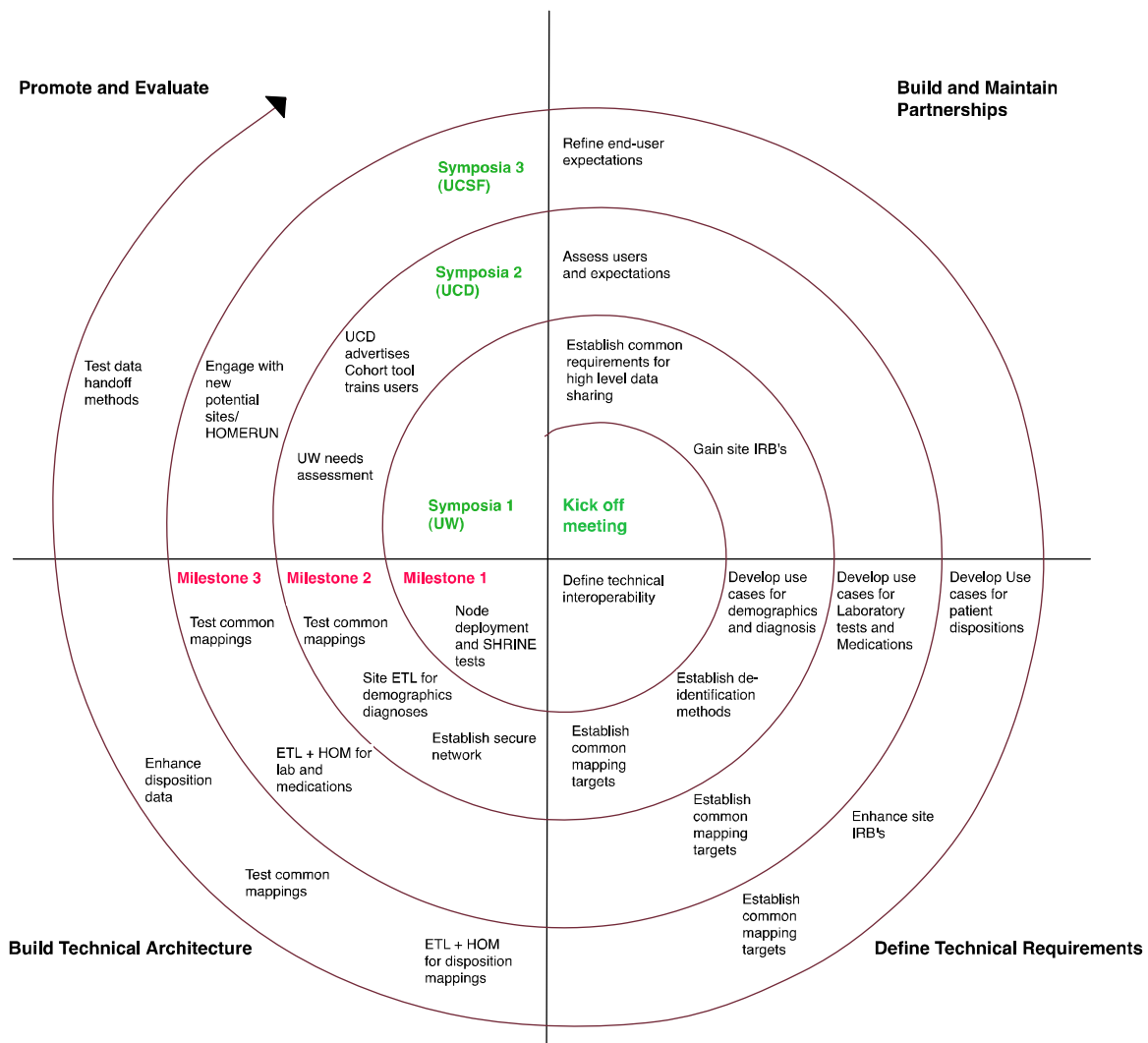


Fig. 3. PCF model applied to CICTR detailing three initial cycles and a fourth rapid sprint cycle to mature the initial launch of the data sharing architecture. Project milestones owed to the funders are highlighted in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

lary list of ordered medications to a navigable (RxNorm-based) terminology tree. This was successful with a very restricted set of medications, but stakeholders had difficulty navigating the results. Further refinement of the use cases resulting from this led to a very restricted set of laboratory tests associated specifically with diabetes diagnosis related groups, which in turn required manual data extraction and mapping at each site. Both these data sources tested the design/build/test partnerships across all sites. Throughout the experience of using the PCF model with CICTR, stakeholder engagement grew critically and was crucial to the design process as the project moved towards launch.

Costs and resources used for the initial pilot build of CICTR included: (1) biomedical informatics personnel at each university site (i.e., faculty lead and programming staff), (2) CTSA subsidized staff resources (i.e., system architects, programmers, and analysts for consultation), and (3) infrastructure costs related to establishing the computing environment at each site and the data loads. The CICTR pilot project was designed and implemented within a 20 month timeframe to comply with funder requirements.

Through partnership in the fourth cycle, the project was adopted as a University of California (UC) wide network, with an additional three sites added and additional support for five years to bridge the five UC health campuses under the UC-Research Exchange (UC-ReX) project. As a direct extension of CICTR, the PCF model of

organization development and management now drives the UC-ReX project through a semantic harmonization lifecycle, and is now capable of providing query access to a population in excess of 14 million patients. CICTR has also been adopted at the University of Washington as a self-service tool (De-Identified Clinical Data Repository, DCDR) broadly offered to researchers engaged with the CTSA who want access to cohort counts of patients.

3.3.2. Cross-cycle observations

By the beginning of the fourth “sustainability” cycle of the project and after completion of project funding, the CICTR project had developed a mature architecture representing four unique data sources on a collective total of over 5 million patient lives seen in three institutions. Throughout the project, the team used regular iteration and feedback with stakeholders and identified and met challenges across themes. Early expectations were set in the project that helped keep partners focused on use cases that would be needed within the promotion and evaluation phases of the project.

But as the resolution of the required data from each site increased, so did the difficulty in developing broad measureable test cases. The cycle process revealed that the project lacked engagement with actual users who sought data and across institutions. Towards the third cycle, the project used the PCF model and themes to assess expansion to additional sites, and as a result

engaged with a new set of comparative effectiveness researchers who explicitly sought multi-site data discovery capabilities. This required developing a strategy to engage the new stakeholders, using reflection on phases and outcomes to date, as well as build common expectations for rapid development and evaluation. This process was captured in documents and expectations for each of the four themes in the spiral, and translated into a two month iterative development sprint that culminated in successful stakeholder-driven use cases across the network.

The application of this iterative approach to growing a network was considered a success due to successful development of a functional data sharing architecture, positive feedback, and continued user engagement. The PCF model also helped accommodate a tightly defined, short timeline driven project scope. The iterative process served to support timely information feedback for all parts of the process, and in turn maintained the project group on common and clear goals.

3.4. Cross-team pilot collaboration

Collaborative meetings were held between Data QUEST and CICTR biomedical informatics team members, sharing progress, technical failures and solutions, ethical considerations, governance documents, and trust building activities. The PCF spiral model was jointly developed by the biomedical informatics teams and adopted to organize and help direct work within the pilot projects, while providing a framework to cross share development activities. In addition to the adoption of the spiral model itself, team members shared governance and ethics work, including specifics for approaches to development of governance infrastructure and supporting documents. Teams also shared technology findings including software, ontology, and data quality experiences. In addition to the process the PCF model offered each project, it also provided a structure for cross team collaboration.

Cross collaboration observations led to the discovery of common challenges and lessons learned across both pilot projects. Both projects were centered in infrastructure based grant proposals with no specific use case driven directive, which created ambiguity for the scope of work. Use cases proved crucial for determining overall data requirements, as well as offering initial test cases to evaluate the system during launch. In hindsight, use cases could have been initially defined in early cycles given both pilot projects required multiple cycles of iteration to define functional use case tests towards the end of the project lifespan. Early definition of use cases may have increased efficiency overall. Both projects also suffered from a lack of users standing in the ready to use the clinical data sharing architectures, creating a need for expanding efforts to promote dissemination of use. Finally, both pilot projects required resource and time intensive effort to create trust between partners, without which the technical architectures could not have succeeded.

4. Discussion

The spiral model offered a practical and flexible process for creating two new electronic health record driven data sharing pilot architectures federated across multiple health care organizations. Creating these complex clinical data sharing networks for research requires a communications-driven process that prioritizes partner engagement throughout all aspects of the project. The cycling supported by the PCF model across these projects ensured technical requirements were iterated closely with partners and facilitated partner engagement in the process. The two CTSA supported pilot project teams also benefited from cross collaboration using the

PCF model, particularly given the similarities in scope of work and complexity of the socio-technical environments.

Both pilot projects were a success in creating functioning data sharing architectures across federated systems, using multiple iterations within a spiral based PCF model of development. The PCF model accounted for key project missions and provided structure, context, and concrete direction for addressing barriers that were often associated with limiting or preventing project success. Success of the PCF model is also evidenced by both project teams opting to add additional cycle iterations to further mature their technical architectures and both projects resulting in successful expansions of the initial projects.

4.1. Lessons learned

Unexpectedly, both projects required additional cycles before sharing actual data across partner data sites. The PCF model allowed for a heuristic evaluation that helped the teams identify inefficient processes and a critical missing component, namely definition of key use cases, which led to the need for additional cycles. Data sharing architectures require clearly defined research or clinical questions and engaged end-users, which are vital for defining system requirements. Both projects were funded specifically to pilot informatics infrastructures with no anchoring specific clinical topic to guide specifications, use cases, and ultimately a user base. As a lesson learned, we identified the crucial need to focus more directly on developing research questions in earlier cycles to frame and provide context for the project. For future architecture builds, creating use cases at the outset of project initiation may limit the need for extra cycles.

Developing data sharing governance with partners impacted system design significantly and should be considered as early as possible. Governance development should include involvement from partners, experienced legal experts, and data extraction stakeholders, and should be revisited during each incremental cycle. Governance stakeholders played a crucial role in system requirement definition from the design, content, and the technical mechanics of how data would be shared, and the methods used to access the resulting data.

Lessons learned highlight the importance of flexibility in implementation management, the on-going complexity of aligning data across each data site, challenges in engaging users, the impact of governance issues on design, and the need to focus on system utility in the early stages of development to sustain development across multi-disciplinary teams. Developing trusted environments is complex and critical for project success and in the cases of these pilot projects, achievable with appropriate resources.

4.2. Limitations

Limitations include the inability to compare the number of cycles and the overall resources and costs needed for these pilot architectures to other similar efforts, given clear baseline data on these metrics with other existing architectures have not been published consistently. However, we were able to use a reasonable number of cycles for each of these pilot projects and stay within resource constraints. We report on only a single approach for developing these architectures, which did not allow for comparisons or testing of different software development models. Future studies could address testing multiple software development models, particularly given convergence and definition of aspects of these models remains an active area of discovery.

Summary points

- Building federated data-sharing architectures requires supporting a range of data owners, effective and validated semantic alignment between data resources, and consistent focus on end-users.
- Establishing these resources requires recognition of development methodologies that support internal validation of data extraction and translation processes, sustaining meaningful partnerships, and delivering clear and measurable system utility.

5. Conclusion

We found that our spiral based PCF model was crucial for creating collaboration between our two pilot projects building functional federated data sharing architectures and provided great utility for promoting success and evaluating challenges within each pilot project. Multiple national efforts have and continue to invest in the development of novel network-based data sharing infrastructure development for research and cross collaborations would strengthen this work and likely increase success. The PCF model may help identify and establish necessary relationships and early detection of barriers within and between teams. Finding cohesive methods that focus on building appropriate early use cases, bringing in users, and systematically building trust among partners are needed to increase implementation success of data sharing architecture development projects.

Future work would benefit from cross collaborations between similar data architecture building projects, definition of use cases early in the process, and proper resources to support work in building trust between partners. Use of software development models can support this future work and help create standard processes for building these complex architectures.

Conflict of interest

We wish to confirm that there are no known conflicts of interest.

Acknowledgements

We would like to thank our Institute for Translational Health Sciences colleagues with the Data QUEST and the CICTR projects and the DARTNet Institute. This research was supported by the National Center for Advancing Translational Sciences, Clinical and Translational Science Award for the Institute for Translational Health Sciences UL1TR000423 and Contract #HHSN268200700031C.

References

- [1] W.G. Van Panhuis, P. Paul, C. Emerson, et al., A systematic review of barriers to data sharing in public health, *BMC Public Health* 14 (2014) 1144, <http://dx.doi.org/10.1186/1471-2458-14-1144>.
- [2] National Institute of Health, NIH Data Sharing Policy, 2003, http://grants.nih.gov/grants/policy/data_sharing/.
- [3] J.M. Westfall, J. Mold, L. Fagnan, Practice-based research—blue highways on the NIH roadmap, *JAMA* 297 (4) (2007) 403–4406.
- [4] C.C. Diamond, F. Mostashari, C. Shirky, Collecting and sharing data for population health: a new paradigm, *Health Aff. (Millwood)* 28 (2) (2009) 454–466.
- [5] A. Wilcox, G. Randhawa, P. Embi, H. Cao, G. Kuperman, Sustainability considerations for health research and analytic data infrastructures, *eGEMS* 2 (2) (2014) (Article 8).
- [6] R. Lazarus, et al., Electronic support for public health: validated case finding and reporting for notifiable diseases using electronic medical data, *J. Am. Med. Inform. Assoc.* 16 (1) (2009) 18–24.
- [7] E.A. Zerhouni, Translational research: moving discovery to practice, *Clin. Pharmacol. Ther.* 81 (1) (2007) 126–128.
- [8] J.S. Ash, N.R. Anderson, P. Tarczy-Hornoch, People and organizational issues in research systems implementation, *J. Am. Med. Inform. Assoc.* 15 (3) (2008) 283–289.
- [9] S.C. Murphy, A. Henry, Security Architecture for Query Tools Used to Access Large Biomedical Databases, American Medical Informatics Association, San Antonio, Texas, 2002.
- [10] S. Murphy, et al., Instrumenting the health care enterprise for discovery research in the genomic era, *Genome Res.* (2009).
- [11] B. Malin, A computational model to protect patient data from location-based re-identification, *Artif. Intell. Med.* 40 (3) (2007) 223–239.
- [12] P. Ohm, Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization, University of Colorado Law Legal Studies Research Paper, 2009, 09–12.
- [13] B. Evans, Congress' new infrastructural model of medical privacy, *Notre Dame L. Rev.* 585 (2009) 619–620.
- [14] N.E. Kass, An ethics framework for public health, *Am. J. Public Health* 91 (11) (2000) 1776–1782.
- [15] H. Nissenbaum, Protecting privacy in an information age: the problem of privacy in public, *Law Philosophy*, 1998, 17, 559–596.
- [16] D.R. Karp, et al., Ethical and practical issues associated with aggregating databases, *PLoS Med.* 5 (9) (2008) pe190.
- [17] S. MacKenzie, M. Wyatt, R. Schuff, J. Tenenbaum, N. Anderson, Practices and perspectives on building integrated data repositories: results from a 2010 CTSA survey, *J. Am. Med. Inform. Assoc.* (2012) e119–e124, 19(e1).
- [18] A. Gupta, et al., Federated access to heterogeneous information resources in the Neuroscience Information Framework (NIF), *Neuroinformatics* 6 (3) (2008) 205–217.
- [19] S.T. Rosenbloom, et al., A model for evaluating interface terminologies, *J. Am. Med. Inform. Assoc.* 15 (1) (2008) 65–76.
- [20] A. Hurd, The federated advantage. Data exchange between healthcare organizations in RHIOs is a hot topic. Can federated models end the debate? *Health Manage. Technol.* 29 (4) (2008) 14–16.
- [21] G.M. Weber, et al., The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories, *J. Am. Med. Inform. Assoc.* 16 (5) (2009) 624–630.
- [22] R.L. Bradshaw, et al., Architecture of a federated query engine for heterogeneous resources, *AMIA Ann. Symp. Proc.* 2009 (2009) 70–74.
- [23] M. McKinney, HIPAA and HIT: tighter control of patient data, *Hosp. Health Netw.* 83 (6) (2009) 50–52.
- [24] C.M. DesRoches, et al., Electronic health records in ambulatory care—a national survey of physicians, *N. Engl. J. Med.* 359 (1) (2008) 50–60.
- [25] P.B. Wise, The meaning of meaningful use. Several technology applications are needed to qualify, *Healthcare Exec.* 25 (3) (2016) 20–21.
- [26] D.W. Bates, A. Bitton, The future of health information technology in the patient-centered medical home, *Health Aff. (Millwood)* 29 (4) (2016) 614–621.
- [27] K.A. Stephens, et al., LC Data QUEST: A technical architecture for community federated clinical data sharing, In: 2012 AMIA Summits Transl Sci Proc., 2012, AMIA: San Francisco, CA.
- [28] N. Anderson, et al., Implementation of a de-identified federated data network for population-based cohort discovery, *J. Am. Med. Inform. Assoc.* 26 (2011).
- [29] A.M. Magdaleno, C.M.L. Werner, R.M. Araujo, Reconciling software development models: a quasi-systematic review, *J. Syst. Software* 85 (2) (2012) 351–369.
- [30] B. Boehm, R. Turner, Using risk to balance agile and plan-driven methods, *Computer* 36 (6) (2003) 57–66.
- [31] B. Boehm, W.J. Hansen, The Spiral Model as a tool for evolutionary acquisition, *J. Defense Software Eng.* 14 (5) (2001) 4–11.
- [32] B.W. Boehm, A spiral model of software development and enhancement, *Computer* 21 (5) (1988) 61–72.
- [33] P. Hitz, B. Johnson, J. Meier, B. Wasbotten, I. Haller, PS3-23: VDW data source: essential health, *Clin. Med. Res.* 11 (3) (2013) 178.
- [34] S. Nagykaldi, et al., Improving collaboration between primary care research networks using access grid technology, *Inform. Prim. Care* 16 (1) (2008) p51–8.
- [35] K.A. Peterson, et al., A model for the electronic support of practice-based research networks, *Ann. Fam. Med.* 10 (6) (2012) 560–567.
- [36] K.A. Stephens, C. Lin, L. Baldwin, A. Echo-Hawk, G. Keppel, A web-based tool for cataloging primary care electronic medical record federated data: FInDIT, 2011, AMIA: Bethesda, MD.
- [37] S.N. Murphy, et al., Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2), *J. Am. Med. Inform. Assoc.* 17 (2) (2010) 124–131.
- [38] G.M. Weber, et al., The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories, *J. Am. Med. Inform. Assoc.* 16 (5) (2009) 624–630.