9-17-2014

Sustainability Considerations for Health Research and Analytic Data Infrastructures

Adam Wilcox
Intermountain Healthcare, adam.wilcox@imail.org

Gurvaneet Randhawa
Agency for Healthcare Research & Quality, Gurvaneet.Randhawa@ahrq.hhs.gov

Peter Embi
The Ohio State University, Peter.Embi@osumc.edu

Hui Cao
Astrazeneca Pharmaceuticals, hui.cao1@astrazeneca.com

See next pages for additional authors

Sustainability Considerations for Health Research and Analytic Data Infrastructures

Abstract

Introduction: The United States has made recent large investments in creating data infrastructures to support the important goals of patient-centered outcomes research (PCOR) and comparative effectiveness research (CER), with still more investment planned. These initial investments, while critical to the creation of the infrastructures, are not expected to sustain them much beyond the initial development. To provide the maximum benefit, the infrastructures need to be sustained through innovative financing models while providing value to PCOR and CER researchers.

Sustainability Factors: Based on our experience with creating flexible sustainability strategies (i.e., strategies that are adaptive to the different characteristics and opportunities of a resource or infrastructure), we define specific factors that are important considerations in developing a sustainability strategy. These factors include assets, expansion, complexity, and stakeholders. Each factor is described, with examples of how it is applied. These factors are dimensions of variation in different resources, to which a sustainability strategy should adapt.

Summary Observations: We also identify specific important considerations for maintaining an infrastructure, so that the long-term intended benefits can be realized. These observations are presented as lessons learned, to be applied to other sustainability efforts. We define the lessons learned, relating them to the defined sustainability factors as interactions between factors.

Conclusion and Next Steps: Using perspectives and experiences from a diverse group of experts, we define broad characteristics of sustainability strategies and important observations, which can vary for different projects. Other descriptions of adaptive, flexible, and successful models of collaboration between stakeholders and data infrastructures can expand this framework by identifying other factors for sustainability, and give more concrete directions on how sustainability can be best achieved.

Acknowledgements

This work was supported in part by R01 HS019853 from AHRQ, Title: Washington Heights/Inwood Informatics Infrastructure for Community-Centered Comparative Effectiveness Research (WICER) and R01 HS022961 from AHRQ, Title: WICER 4 U.

Keywords

sustainability, data reuse, health information technology

Disciplines

Health Information Technology | Health Services Research

Authors
Adam Wilcox, *Intermountain Healthcare*; Gurvaneet Randhawa, *Agency for Healthcare Research & Quality*; Peter Embi, *The Ohio State University*; Hui Cao, *Astrazeneca Pharmaceuticals*; Gil Kuperman, *NewYork-Presbyterian Hospital*.

Sustainability Considerations for Health Research and Analytic Data Infrastructures

Adam Wilcox, PhD;[i] Gurvaneet Randhawa, MD, MPH;[ii] Peter Embi, MD, MS, FACP, FACMI;[iii] Hui Cao, MD, PhD, MS;[iv] Gilad J. Kuperman, MD, PhD, FACMI[v]

Abstract

Introduction: The United States has made recent large investments in creating data infrastructures to support the important goals of patient-centered outcomes research (PCOR) and comparative effectiveness research (CER), with still more investment planned. These initial investments, while critical to the creation of the infrastructures, are not expected to sustain them much beyond the initial development. To provide the maximum benefit, the infrastructures need to be sustained through innovative financing models while providing value to PCOR and CER researchers.

Sustainability Factors: Based on our experience with creating flexible sustainability strategies (i.e., strategies that are adaptive to the different characteristics and opportunities of a resource or infrastructure), we define specific factors that are important considerations in developing a sustainability strategy. These factors include assets, expansion, complexity, and stakeholders. Each factor is described, with examples of how it is applied. These factors are dimensions of variation in different resources, to which a sustainability strategy should adapt.

Summary Observations: We also identify specific important considerations for maintaining an infrastructure, so that the long-term intended benefits can be realized. These observations are presented as lessons learned, to be applied to other sustainability efforts. We define the lessons learned, relating them to the defined sustainability factors as interactions between factors.

Conclusion and Next Steps: Using perspectives and experiences from a diverse group of experts, we define broad characteristics of sustainability strategies and important observations, which can vary for different projects. Other descriptions of adaptive, flexible, and successful models of collaboration between stakeholders and data infrastructures can expand this framework by identifying other factors for sustainability, and give more concrete directions on how sustainability can be best achieved.

Introduction

The increasing prevalence of electronic health record (EHR) systems and other health-related data sources is enabling the development of data infrastructures that can more efficiently support research needs in health care, including patient-centered outcomes research (PCOR) and comparative effectiveness research (CER), as well as a wide variety of biomedical research and health delivery-related operational questions. PCOR and CER are seen as emerging approaches to efficiently identify ways to improve health care delivery.[1] To address challenges to the use of health-related data for research, the Agency for Healthcare Research and Quality (AHRQ) has funded several large projects to develop an infrastructure to both support and demonstrate the value of CER and PCOR,[2] as well as to increase the understanding of issues related to creating and using the data infrastructures. The Electronic Data Methods (EDM) Forum, also an AHRQ-funded program, has collaborated with the infrastructure projects to facilitate cross-project learning and to synthesize and disseminate lessons learned.[3–5]

Among the most important issues that have been recognized and considered by the AHRQ-funded projects and the EDM Forum is how to sustain the research data infrastructure beyond the initial investment. The investment in the large infrastructure projects and the EDM Forum was funded through the American Reinvestment and Recover Act of 2009.[3,5] This funding was expected to be a one-time instance, lasting only three years during the initial development stage, though the projects themselves were expected to last beyond the development funding. As a result, efforts to define challenges and solutions to data infrastructure sustainability have been substantial over the course of the projects. In this paper, we have convened a broad group of experts who have participated in these defining efforts. The experts include a principal investigator of an AHRQ-funded infrastructure project, the program officer for the group of projects, a clinical research informatics expert actively applying a sustainability strategy for an existing institutional infrastructure, an informatics researcher working in the pharmaceutical industry, and the leader of a sustained health information

[i]Intermountain Healthcare, [ii]Agency for Healthcare Research & Quality, [iii]The Ohio State University, [iv]Astrazeneca Pharmaceuticals, [v]NewYork-Presbyterian Hospital

exchange infrastructure. We present perspectives on and lessons learned about sustainability based on the authors' experiences and anecdotal observations facilitated by the EDM Forum. We first discuss sustainability in the context of data infrastructure platforms and factors important for sustainability. These factors (assets, expansion, complexity, and stakeholders) were identified through iterative discussions among the authors as dimensions where different projects and sustainability strategies varied, and thus would be critical to evaluate when developing and assessing strategies. For example, projects with different assets, expansion goals for those assets, complexity of the assets, and available stakeholders will have different sustainability plans, which must address each of these factors. We then enumerate important lessons we have learned regarding infrastructure sustainability, relating these lessons back to the important sustainability factors. These lessons can provide basic guidance in creating viable data infrastructures for research.

Sustainability Factors: Assets, Expansion, Complexity, and Stakeholders

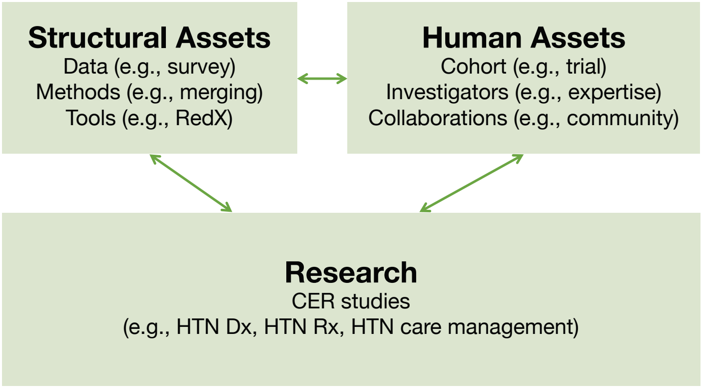
Researchers have considered issues of sustainability in related domains, such as quality improvement,[6–8] public health,[9,10] and biological databases.[11] In each of these areas, they have identified the elements that needed to be sustained and factors that influence sustainability. Edwards et al. created a framework for sustaining health care safety and quality projects that considers both the goals for sustainability and a model of knowledge dissemination.[6] Schell et al. created a framework of public-health program capacity sustainability that included nine related domains.[10] And Chandras et al. reviewed different models that had been employed in financial sustainability and preservation of large biological databases.[11] Each framework differed substantially due to differences in the domains for the frameworks and their goals.

A health research data infrastructure is complex and multifaceted, and its sustainability framework will individually differ from frameworks in other domains. It includes technical, organizational, and policy components, and its sustenance requires staff and nonpersonnel resources. With one data infrastructure project, researchers specifically identified the different components that were part of the sustainability framework (see Figure 1). While some components of the project—such as data, methods and investigators—would be similar for other data infrastructure projects, other components such as the patient cohort and tools are specific to the project. In general, though, the categories of components, or *assets*, are consistent across projects.

Assets

The infrastructure includes operational informatics components that have direct financial costs. In Figure 1, these are identified as "Structural Assets." Technical resources include (1) servers, which have hardware and software maintenance expenses; (2) data interfaces, which may have been developed as part of an initial investment but must be maintained over time; and (3) application software, such as analytics software, which will need to be

Figure 1. Initial Sustainability Framework from the WICER Project[6]



Notes: Some assets would be similar across other data infrastructure projects, but others are specific to the project. Different assets are interrelated components of the larger project.

maintained and may need to be extended over time to serve new purposes. Staff support is also needed to support various aspects of the infrastructure. While technical support may be outsourced, staff needed to support the research itself (e.g., lead the scientific projects, extract data, perform analyses, etc.) are defined by the specific project. Staff are also needed for overall administration, project management, stakeholder relationship management, communications, grant writing, legal activities, etc. Some level of technical and organizational staffing needs uninterrupted support to prevent decay, while others are flexible according to the research project need. For example, health data without sustained administration for storage and security should be eliminated to prevent loss through security breaches.

A health research data infrastructure also includes "in-kind" activities by the participants that may not have direct financial costs but require operational staff support. These are some of the "Human Assets" of the infrastructure, and are based on activities and relationships among people. Examples of such activities include (1) development and support of governance structures; (2) development of collaborative networks, including the creation, maintenance, and extension of stakeholder engagement; (3) policy development, e.g., privacy; and (4) strategic planning. These activities may entail nontrivial organizational work. Among the AHRQ-funded infrastructure projects, these components were often the most overlooked components of the project in initial design, but the most difficult to develop. For example, data governance was noted as a consistent issue that was time- and labor intensive.

Lastly, the infrastructure includes scientific components that need to be sustained. They include such human assets as (1) scientists and their associated expertise, (2) the research subjects or the cohorts that have been developed over time, and (3) the relationships among the research team that have been developed to enable the scientific products to be created. The scientific components also include such data assets as databases, scientific techniques, and tools. In Figure 1, these are mainly the "Research" assets, though they include project-specific Human Assets.

Identification of assets is critical in the development of a sustainability plan. Failure to identify the core assets of a project and their necessity to the project can undermine a sustainability strategy. If only the technical assets are supported and the human assets are ignored, the system may be operationally supported but decay in use over time.

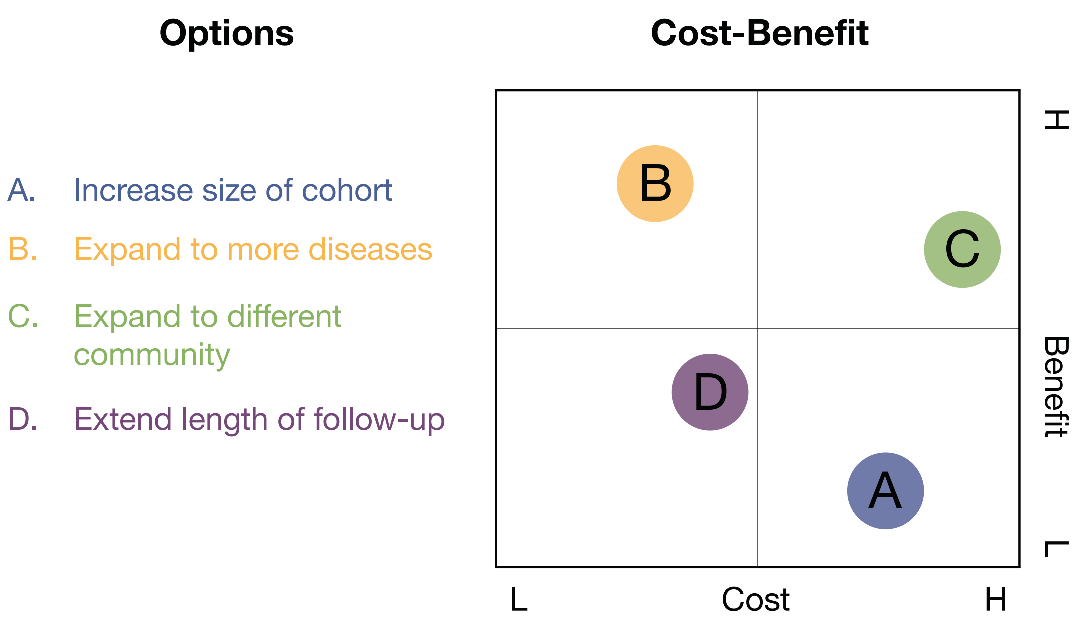
Expansion

For a research infrastructure to be sustained, all components—technical (informatics), governance, scientific, etc.—must be sustained. However, the different components of the infrastructure may have different sustainability requirements. For example, a project may have informatics components that are easily sustained because they require minimal ongoing operations costs. The same project may have complex, evolving governance issues as the infrastructure expands to include other participants. Sustainability considerations change according to the level of expansion needed to keep it functional.

In some instances, sustainability may mean maintenance at the same base level of functionality or capacity as when the project was initially created, with no expansion. This may occur with a data set from a cross-sectional survey, where the main investment initially was the data collection, but once the data have been collected the only support needed is to sustain a capability that enables other researchers to access it for answering research questions. Another low-maintenance example is an infrastructure to sustain a single research study or group of studies. Such projects are usually smaller, have infrastructure that is tightly matched to the specific research focus, and may not be generalizable or malleable to other topics or approaches. In contrast, in other instances, a project's sustainability may mean some level of growth to accommodate an expansion of the initial area of focus, new areas of focus, and new stakeholders. Such projects require some amount of continued investment to expand the overall infrastructure—e.g., to support the inclusion of additional data feeds or additional kinds of analytic software—as well as operational maintenance and support of the platform.

A specific example of how expansion related to sustainability was from the Washington Heights/Inwood Informatics Infrastructure for Comparative Effectiveness Research (WICER), initially led by Dr. Wilcox. This project included a community survey of the local population, who were then a cohort for further studies. Without some expansion of the community survey, the data would eventually become stale and outdated, so investigators considered different options for extending the survey, leveraging the infrastructure that had already been created to collect it. The original cohort consisted of about 6,000 members of the Washington Heights/Inwood community, who were surveyed at two different times regarding overall health status, perceptions, and characteristics. Specific attention was also given to measures relating to hypertension. Four expansion options were considered, based on extending the survey content or the cohort. These options included increasing the size of the cohort within the same community, creating a new cohort in a different community, expanding the content focus to additional diseases (the initial focus was on hypertension), or increasing the number of successive surveys to make it more longitudinal. A simple cost-benefit comparison was made by ranking the different options in terms of cost and benefit, with the best approach identified as expanding the survey to different diseases using the same cohort (Figure 2).[6] This deliberate consideration of options for expansion was critical to the WICER sustainability strategy.

Figure 2. Cost-benefit comparison of expansion options for the WICER cohort



While the examples above describe the levels of expansion generally for an infrastructure, the levels actually apply more directly to each element. That is, even with an infrastructure that must be expanded to be sustained, there may be some components of that infrastructure that only need to be maintained at current levels of functionality (e.g., tools for navigating data may not need to expand when other research areas are added to the overall infrastructure). Even among components that expand, the rates of expansion may be different. A distributed research network may expand linearly in terms of data as new sites are added, but the costs of responding to ongoing queries may be higher as researchers at a new site request data across the full network.

Complexity

Complexity is a third factor related to sustainability. In addition to the amount of resources needed to sustain a particular component of the infrastructure, some components are more complex than others to sustain. For example, informatics assets are often well-defined, with well-understood ongoing operations costs and implementation methods. Expansion of the technology or the technical support usually translates to a linear expansion in the costs of the hardware, software, or support staff. On the other hand, other components may have a much higher level of complexity, and do not map directly to costs. Governance and collaborations are difficult to define ongoing operational costs for, because they are dependent on a multitude of other factors that are often not easily measured and, therefore, require more constant navigation to sustain. Scientific components can be more or less complex, depending on the specific application and its existing support in the infrastructure.

Complexity affects a sustainability strategy most directly by affecting the resources required to sustain an asset. Usually these resources are costs, though as discussed above, they can also be other investments that do not translate directly to costs such as attention, effort, and political capital.

Stakeholders

So far, we have discussed considerations of sustainability related to the characteristics of the data infrastructure itself. But the value of a resource is critical to sustainability, and value is determined by groups external to the resource. The diversity of stakeholders and of their needs means that it will be used for different purposes, and different value propositions will influence how the infrastructure is sustained (see Table 1). The infrastructure can be used for multiple purposes including research (e.g., CER, PCOR, or to better understand the natural course of a disease), surveillance (e.g., drug and device safety outcomes), improving quality of care and tracking quality metrics (e.g., by sharing best practices and providing benchmarks), improving clinical operations and efficiency of a health care delivery organization, improving a clinician's workflow, providing convenience to patients in their interactions with health systems, and conducting market research. These uses are of potential interest to many stakeholders. These include government (federal, state, and local), nonprofit research funding organizations (e.g., PCORI, foundations), for-profit organizations (e.g., pharmaceutical and device-, information technology-, and telecommunications industries), employers and insurers, health care delivery organizations, and individuals (e.g., clinicians and patients).

Table 1. Examples of Uses and Stakeholders for a Research Data Infrastructure

Uses	Stakeholders
Research	Government
Surveillance	Nonprofit research funding organizations
Tracking quality metrics	For-profit companies
Improve clinical operations	Employers, insurers, and other payers
Patient access	Health care delivery organizations
Market research	Individual clinicians and patients

Notes: The examples for each type indicate the broad diversity of stakeholders and their needs. The lists are not exhaustive, but are indicators of the variation.

A stakeholder may have multiple needs that can be met by the infrastructure. For example, a federal research agency may value knowing the outcomes of a patient transitioning from an inpatient facility to a long-term care facility and would fund this research or it may be interested in improving quality of care or health system efficiency and would be willing to support such activity. Another federal agency may be interested in conducting post-market surveillance to learn safety outcomes of a newly introduced drug in diverse patient groups treated in different health care settings as part of its regulatory function, or it may be interested in the use of the infrastructure for trials done as part of a premarket approval process for a new drug. A for-profit company that makes medical devices may need information from a network for market research, for conducting a new trial, or for monitoring safety outcomes. A clinician may participate in a network to learn best practices from others or to gain new capabilities in efficiently managing his or her panel of patients. A patient may be willing to pay subscription fees to use a mobile application: to keep track of test results, to monitor progress toward a goal, or for access to a clinician to avoid an in-person visit when it is clinically prudent to do so. Even researchers, for whom the infrastructure was initially developed, can have differing needs of the infrastructure. For example, researchers focusing on the comparative effectiveness of different treatments for a particular disease may require very detailed symptom and treatment data about patients that have that disease. In contrast, a health care provider organization that is using an analytics infrastructure to support a broad portfolio of quality improvement projects may require data on a broader set of patients but may not require the same detailed depth of data as do the disease-focused researchers.

Stakeholders are critical to sustainability because they can provide resources to a data infrastructure or other system in exchange for its use. If value and use can be defined for stakeholders, their contributions to the infrastructure can be recommended. These contributions may be financial (such as grant awards or subscription fees) or in-kind (e.g., personnel or data to support or enhance the data infrastructure). The value that stakeholders can provide often varies according to their engagement in the project, both in their ability to contribute financially and toward the progress of the infrastructure. The WICER project specifically measured the effect of engagement, or "enthusiasm," of different participants in the research infrastructure.[6] Two participating organizations each had similar activities for contributing data to the infrastructure (Figure 3). One organization was more engaged in the overall goals of the project, and was able to easily accomplish the tasks needed to contribute data, such as reaching consensus on the data to be contributed, submitting an initial data set, and providing updates to that data. Since this was done early, the remaining work with the organization was refining the data and needs to support the shared vision of the project. The second organization was less engaged in the goals of the project. This led to much greater effort being required to both reach consensus and provide an initial data set. In the end, we estimated that the tasks completed by the enthusiastic partner were accomplished with about one-third of the effort compared with the effort of the reluctant partner. This experience was important for the project in the development of its sustainability plans.

Summary Observations

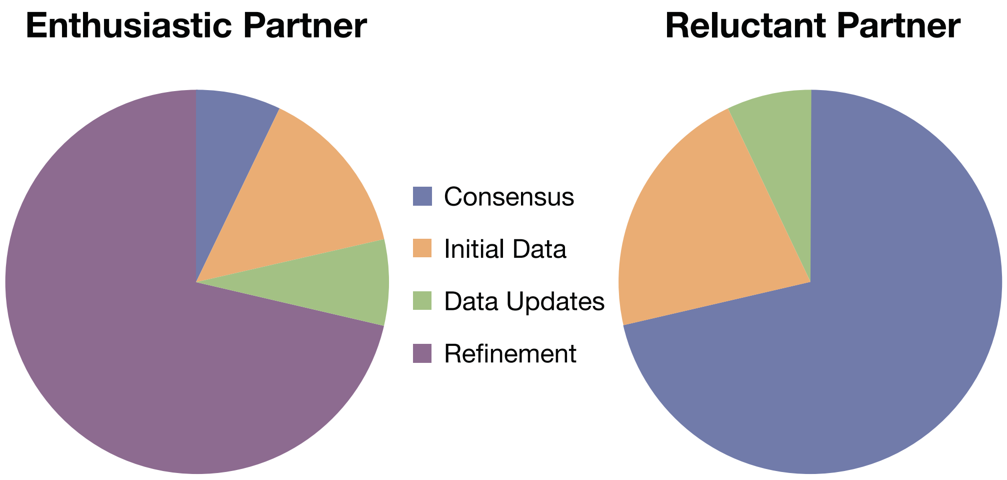
The four factors or dimensions listed above—the type of asset, the level of expansion, its complexity, and stakeholders—are critical in considering a strategy for sustaining a health research and analytic infrastructure. For each component of the platform, the expansion level, complexity, and existing or potential stakeholders will define appropriate approaches for sustaining it, and will inform an overall strategy. The interaction among the factors is important for the sustainability plan. Based on our observations of the EDM projects and similar projects that considered sustainability, we have identified some general findings about the sustainability factors and their interactions. These findings can be useful for understanding the relationships between the factors, and for identifying barriers and opportunities relevant to a sustainability approach.

Lesson 1: There Are Significant Ongoing Costs in Supporting a Research and Analytic Infrastructure

The costs of sustainability are significant and can easily be underestimated. Industry's annual maintenance costs for software and hardware are generally 20 percent of the purchase costs, but with an infrastructure, it can exceed that number, especially where the sustainability involves some extension of the architecture. Ongoing use of data usually involves additional data transformations for successive projects, which require expertise in both understanding and representing the data. In one of the AHRQ projects that aimed to create a community registry that included patient-supplied data, project investigators estimated the annual ongoing maintenance costs for the research data warehouse and data exploration tools to be about 35–40 percent of the initial development costs.[6] About half of these costs were personnel costs for data analysts, who would query and transform the data for successive projects (Table 2). Other costs included leadership (project and team management, as well as resource advocacy), database hardware and software, database administration, and application development.

Chandras et al. found the informatics component alone to be a significant part of the overall project funding for a biological database, about one-fifth of the overall development.[7] In the projects reviewed by the EDM Forum, there were significant costs in terms of leadership and effort needed just to maintain the data sharing partnerships for a research data infrastructure.[3]

Figure 3. Comparison of Effort Required and Results Achieved for Partners and Stakeholders with the WICER Project



Notes: The enthusiastic partner overcame barriers more easily, and accomplished tasks similar to those of the more reluctant partner—but with about one-third of the effort.

Table 2. Development and Maintenance Costs for a Research and Analytic Data Infrastructure

Resource	Startup Costs	Annual Costs	5-year Costs (development)	% of Total	% of Annual
Leadership	—	$140K	$700K	21%	28%
Database	$350K	$20K	$450K	14%	4%
Database administration	—	$63K	$315K	9%	12%
Data analysts	—	$225K	$1125K	34%	45%
Development	$450K	$56K	$730K	22%	11%

Notes: The major cost for maintenance was for data analysts, who would query and transform the data for successive projects. These costs would be expected to increase linearly as the project proceeds, unless tools are implemented to increase self-service use of the data infrastructure.

This lesson is a relationship between the assets, their expansion needs, and their complexity. The costs are substantial because different assets can require expansion to be effectively sustained, and the expansion of the assets can increase their complexity and costs.

Lesson 2: Costs Can Be Covered if Value Can Be Created

A challenge in sustaining an infrastructure is that the continued support will be based on the ability of the infrastructure to demonstrate value.[8] While the initial investment might be justified on the potential of the benefits, continued support for maintenance and expansion is not likely without first receiving some benefit. This lesson is a direct relationship between assets and stakeholders—the value of the assets is defined by their value to stakeholders, who can then determine whether to participate in sustaining the assets. Support for ongoing operations costs likely will have to come from the stakeholders who are directly receiving that benefit, rather than stakeholders in the initial capital investment stage, whose goal may have been simply to establish the infrastructure. An example is a research infrastructure that may be initially developed with a capital investment from public funds, where the public benefits generally from improved research capability. After the infrastructure is developed, however, the maintenance costs may need to shift to the researchers who are actually using the infrastructure directly. The researchers may not have the capability to fund the development, but would be able to contribute to the maintenance once the infrastructure is sufficient to directly benefit their research projects.

For-profit companies could also contribute to maintenance costs after the infrastructure is developed. Well-designed health research and analytic data infrastructures are valuable for the biopharmaceutical industry, and notable investments have been made to support them. The Observational Medical Outcomes Partnership (OMOP) established a network of data sources for research,[9] funded by a consortium of pharmaceutical companies. Other examples include the Merk-Moffitt,[10] Merck Regenstrief,[11] and Intermountain Deloitte collaborations. Longitudinal patient phenotype data, especially with linked genomic data, can also allow companies to gain a better understanding of unmet medical needs, heterogeneity and complexity of chronic diseases, and outcomes of existing treatments. This could then accelerate research and development innovations in the life sciences industry, providing additional value and impact from the created data infrastructure.

There may be some projects, or components of a project, where value for stakeholders cannot be created. These are important to discover so that investments can be made elsewhere. Not every project that is created should be sustained.

Lesson 3: Having Multiple Stakeholders Increases the Opportunity and Complexity of Sustaining an Infrastructure

In general, individual stakeholders have been unable to create and sustain data infrastructures independently, so any long-term sustainability strategy must include and address the needs of multiple stakeholders. The biggest benefit to multiple stakeholders is that maintenance costs are shared among multiple participants. Sometimes cost sharing is done through partnerships, such as academic-commercial partnerships.[7] In health care, these are common in academic medical centers, where partnerships are established between research institutions and health care providers. In addition to the cost sharing benefits, there are other effects from the pairing, such as changes to the infrastructure or its use. For efficiency, the infrastructure tends to adapt so that it focuses on areas of common interest, or so that one stakeholder may flexibly adapt their use of the infrastructure to common areas of other stakeholders.

With a health research data infrastructure, cost sharing commonly occurs with academic research and quality improvement initiatives using data from the same infrastructure, especially with translational research that may have a high overlap in data needs with quality improvement. Research funding cannot sustain the costs of the infrastructure alone, and a quality improvement effort may not be well positioned to gather the same data as effectively as a research effort would. Quality improvement can benefit from the rigorous techniques used to validate the data for research, and the value obtained by the quality improvement may incent the care organization to participate in sustaining the infrastructure. This symbiosis has led research informatics leaders to explicitly target quality-improvement initiatives as opportunities to sustain and support research data infrastructures, and to promote research activities that best align with these efforts.[12]

A challenge, however, is that the multiple stakeholders will also increase the complexity and governance of the infrastructure. That is, the different stakeholders can affect the complexity of the assets. Generally, a research data infrastructure will need to expand over time, and needs for expansion may differ when there are multiple stakeholders. These differences in priorities arise from reasonable differences among stakeholders in such factors as relevant cohorts, time horizons or urgencies of data needs, areas of focus, and perceptions of the ultimate goal for the infrastructure. One example is when data are needed from EHRs. A provider organization may be seeking to use the data to support efficient clinical care, whereas researchers may be seeking to develop a disease-specific research registry. There may be a set of data that are useful to both stakeholders, such as demographics, encounters, ancillary reports, problems, medications, allergies, and financial claims data. However, there may be additional data that are desired by the provider organization to support its goals and a different set of data necessary to support the registry.

Lesson 4: Stakeholders May Support an Infrastructure to Keep Flexibility in an Emerging Area

Sometimes the value provided from a data infrastructure is its potential to expand, rather than its direct use. Organizations that supply data to the research and analytic infrastructure may contribute to ongoing operations costs based on potential rather than immediate value. For example, a health care institution may contribute resources to sustain the project after the initial development even when it doesn't meet an immediate need, because they feel it is more cost-effective than rebuilding the project later. Such institutions may be unlikely to pay for the initial development of the infrastructure due to competing institutional priorities, but the institution has an incentive to maintain the infrastructure once the infrastructure is created, rather than have to fully fund its development (or redevelopment) later. Another reason the institution may support the maintenance is that the analytic infrastructure, once developed, supports the same functions that are requested of the institution, and the existence of a method to support the functions increases their urgency among stakeholders. This was the experience of one of the AHRQ projects, where the tools that were developed as part of the research infrastructure were also seen to be useful to the analytic needs of one of the participating institutions. As a result, the institution contributed directly to the maintenance (and expansion) of that resource.

This lesson relates the expansion of the assets to the stakeholders, who sustain them for their ability to expand. The real value of an electronic data infrastructure is elusive to demonstrate, but the potential remains high. The amount of electronic data available is increasing beyond our ability to manage it, which both decreases our ability to use it fully and increases its potential value. Our experience is that this rapid expansion of electronic data, either through the increased collection of information in electronic form or the integration of data from different sources, have made creating an infrastructure more difficult and more urgent. Data integration itself is challenging, both in the governance among data contributors and the data modeling required to merge the data. These can be done more efficiently as a research initiative, where contributors may participate with the goal of promoting a general good, rather than allowing an advantage for a competitor. The merging of data then greatly increases the availability of data to analyze by contributors. The result is that stakeholders are now challenged with consuming this data effectively, but are not able to consume the data at near the level it is made available. This may lead to concern by stakeholders that they are falling behind other organizations that may be investing in data consumption differently, or just missing opportunities to leverage the data effectively. Broader prominent trends and statements of potential value in the use of data, such as "Big Data," increase this concern. Making sense of the data is thus something desired by many, and while a research data infrastructure may not address all the data consumption needs of a stakeholder, it moves in that direction.

Conclusion

Sustained research data infrastructures will be necessary to support the ongoing health research needs of the country. For example, it is needed to efficiently support the increase in PCOR and CER to inform health care reform. Through the AHRQ, the United States has made an initial investment in creating data infrastructures to help meet that need, but government research grant funding alone cannot and should not fully sustain them. Flexible sustainability must thus be a priority for research data infrastructure projects. We have defined a broad scope of how sustainability strategies can vary for different projects, using perspectives and experiences from a diverse group of experts. This variation is substantial, such that aggregate statistics among projects are difficult to interpret. While some concrete examples have been given, these examples are limited to a few specific projects, and other projects would not be expected to be identical. However, by linking the lessons learned to relationships between the broader factors for sustainability, we believe our conclusions are generalized enough to be applicable to most variations. We propose that others describe adaptive, flexible, and successful models of collaboration between stakeholders and data infrastructures that have, and can continue to, sustained them, identifying other factors for sustainability. As a whole, these case studies could help expand this framework and identify patterns for sustainability, such that more concrete directions on achieving sustainability for a project can be followed.

Acknowledgements

This work was supported in part by R01 HS019853 from AHRQ, Title: Washington Heights/Inwood Informatics Infrastructure for Community-Centered Comparative Effectiveness Research (WICER) and R01 HS022961 from AHRQ, Title: WICER 4 U.

References

1. Sullivan P, Goldmann D. The promise of comparative effectiveness research. JAMA J Am Med Assoc. 2011 Jan 26;305(4):400–1.
2. McKinney M. "Huge potential" for EHRs and comparative effectiveness. Hosp Health Netw AHA. 2010 Apr;84(4):41–2.
3. Holve E, Segal C, Lopez MH, Rein A, Johnson BH. The Electronic Data Methods (EDM) forum for comparative effectiveness research (CER). Med Care. 2012 Jul;50 Suppl:S7–10.
4. Johnson BH. Electronic Data Methods (EDM) forum building the infrastructure to conduct comparative effectiveness research and patient-centered outcomes research using electronic clinical data. Med Care. 2012 Jul;50 Suppl:ii.
5. EDM Forum. Building the Electronic Clinical Data Infrastructure to Improve Patient Outcomes: CER Project Profiles. [Internet]. AcademyHealth; 2012. Available from: http://repository.academyhealth.org/edm_briefs/7/.
6. Wilcox A, Yoon S, Boden-Albala B, Bigger JT, Feldman PH, Weng C, et al. Developing a Framework for Sustaining Multi-institutional Interdisciplinary Community Participatory Comparative Effectiveness Research. AMIA 2013 Joint Summits on Translational Science. San Francisco, CA; 2013.
7. Chandras C, Weaver T, Zouberakis M, Smedley D, Schughart K, Rosenthal N, et al. Models for financial sustainability of biological databases and resources. Database J Biol Databases Curation. 2009;2009:bap017.
8. Thornewill J, Dowling AF, Cox BA, Esterhay RJ. Information infrastructure for consumer health: a health information exchange stakeholder study. Am J Prev Med. 2011 May;40(5 Suppl 2):S123–133.
9. Ryan PB, Madigan D, Stang PE, Overhage JM, Racoosin JA, Hartzema AG. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Stat Med. 2012 Dec 30;31(30):4401–15.
10. Merck-Moffitt Partnership Breaks Down Silos - Bio-IT World [Internet]. [cited 2013 Jun 13]. Available from: http://www.bio-itworld.com/BioIT_Article.aspx?id=78252
11. Regenstrief Institute, Merck partner for personalized care | Healthcare IT News [Internet]. [cited 2013 Jun 13]. Available from: http://www.healthcareitnews.com/news/regenstrief-institute-merck-partner-personalized-care
12. Embi PJ, Kaufman SE, Payne PRO. Biomedical informatics and outcomes research: enabling knowledge-driven health care. Circulation. 2009 Dec 8;120(23):2393–9.