Published OnlineFirst August 28, 2019; DOI: 10.1158/1055-9965.EPI-19-0043

Commentary

Cancer Epidemiology, Biomarkers & Prevention

Addressing Challenges in Converting Grant-Funded Infrastructures to Broadly Used Research Resources

Betsy Rolland[1] and Ann M. Geiger[2]

Abstract

The NCI invests heavily in research resources to serve the research community, including datasets, biospecimen banks, and networks of institutions in which clinical trials and other human subjects research are conducted. These resources often begin as grant-funded infrastructure initiated by scientists based on their own scientific interests, with a subsequent recognition of additional scientific uses. Although converting existing project-specific research activities into research resources may appear efficient in terms of time and financial investment, challenges can arise that undermine this efficiency and jeopardize future use. Here, we describe three challenges in the conversion process: (i) project-based infrastructure versus a research resource for a broader research community; (ii) complexity versus ease of use; and (iii) individual professional goals versus research resource priorities. We use our experience with the NCI-funded Cancer Research Network, particularly the Virtual Data Warehouse, to illustrate each challenge, concluding with strategies to mitigate each one. As studies grow in size and complexity, an ever-increasing volume of data, biospecimens, and human subjects research networks will be available for conversion to resources for scientific questions beyond those originally proposed. Addressing likely challenges thoughtfully can result in a more efficient conversion process and ultimately greater scientific impact.

Introduction

Over the past two decades, the NCI has funded many large research networks charged with addressing complex scientific questions, including the Early Detection Research Network, the Breast Cancer Surveillance Consortium, and the Cancer Research Network (CRN). The overarching aim of such projects is to answer specific research questions, but the institutional infrastructure, scientist partnerships, data, biospecimens, and other potentially reusable research products that develop during the projects may be useful in answering additional questions. In theory, this reuse should accelerate the pace of scientific discovery and create efficiencies by avoiding duplication of costly infrastructure-building efforts. For example, a group of scientists who have been working together in a partnership around one topic can be tapped to explore a related topic, taking advantage of the existing relationships and coordination mechanisms, with lower start-up costs than a new network. While a project-specific research resource may morph organically into a reusable research resource, others are explicitly targeted for a more intentional, deliberate conversion process.

The key steps involved in converting research infrastructure created to answer specific questions into a resource with multiple potential uses are straightforward. First, scientists and their staff build the infrastructure necessary to accomplish the scientific goals of their grant. Second, the infrastructure grows and matures to the point where others see it as a potential resource for answering other scientific questions. Finally, standardized approaches and policies for use of the infrastructure are developed and shared with the scientific community, who can now access a new research resource. The required involvement and cooperation of numerous scientists in this process can create tensions with the potential to undermine the availability and benefits of the resource. For example, delays may result in data sets that are outdated or in staff departures that diminish network capacity.

In this article, we identify three common tensions that can jeopardize the intentional conversion of grant-funded infrastructures to research resources. We illustrate these tensions with examples from our experience working with the CRN and then suggest strategies to mitigate these tensions. Thoughtful planning for these tensions is essential to timely creation of widely accessible, readily usable, and highly impactful research resources from grant-funded activities.

Example of the CRN

The CRN is a collaboration of integrated delivery systems with formal research programs funded through federal and nonfederal grants and contracts. The CRN received three cycles of NCI funding that focused on the development of research infrastructure to support scientist-initiated, large-scale research projects on par with the NIH R01 mechanism (1, 2). After an external evaluation called for increased involvement of scientists outside the CRN to ensure the infrastructure be used for maximum scientific impact, NCI funded a fourth cycle focused primarily

[1]Carbone Cancer Center and Institute for Clinical and Translational Research, University of Wisconsin-Madison, Madison, Wisconsin. [2]Cancer Care Delivery Research, National Cancer Institute Community Oncology Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, Maryland.

Corresponding Author: Betsy Rolland, University of Wisconsin-Madison, Madison, WI 53705. Phone: 608-265-3185; Fax: 608-263-8613; E-mail: brolland@wisc.edu

Cancer Epidemiol Biomarkers Prev 2019;28:1559–62

doi: 10.1158/1055-9965.EPI-19-0043

Published OnlineFirst August 28, 2019; DOI: 10.1158/1055-9965.EPI-19-0043

on the goal of converting the existing infrastructure to a research resource for the broader cancer research community, eliminating funding for specific research projects. A specific requirement was that the network-wide data infrastructure, the Virtual Data Warehouse (VDW), be more accessible to qualified scientists seeking to answer potentially high-impact scientific questions (3).

Several challenges emerged as the CRN moved away from its original research focus to serving as a data resource. Many of these challenges stemmed from the lack of data interoperability across the participating delivery systems or from within delivery systems themselves as information technology platforms changed over time. The quality of available administrative and clinical data also varied depending upon the research question, delivery system, data source, and time frame, requiring a substantial investment of time and programming resources to generate research-quality data for each new study. Scientists working within the CRN were aware of these challenges and able to navigate the systems productively throughout earlier iterations of the Network; however, non-CRN scientists struggled to understand that it was not possible to request and receive a self-contained dataset within a period of days or weeks, as they could with other, more standard available datasets. Instead, non-CRN scientists needed to work closely with a CRN scientist, often one at each participating CRN system, to explore the availability and quality of data pertinent to their research question, requiring a large commitment of time and resources to accomplish their goals. Furthermore, CRN scientists found that providing support to non-CRN scientists did not typically align with their own research interests and professional goals, plus required a time investment that limited CRN scientists' ability to produce publications and grant applications using the data resource they had created. Because involvement of non-CRN scientists in the design of the system was limited, these issues did not surface until after the system had already been developed, making fixing the issues more difficult.

Challenges in the Conversion Process

The challenges encountered are not unique to the CRN and are common barriers to the conversion of infrastructure created for specific research projects into a widely available research resource with many potential uses. Drawing upon our experience designing, developing, and evaluating research resources, here we describe three major challenges inherent to the conversion process: (i) project-based infrastructure versus a research resource for a broader research community; (ii) complexity versus ease of use; and (iii) individual professional goals versus research resource priorities. While the specific details of each challenge may look different in every network, there are common challenges across networks and conversions that we highlight here.

Project-based research infrastructure versus a research resource for a broader research community

Project-based research infrastructure is designed and developed with an overarching goal, a targeted end user, and a specific set of research questions to address (4). However, designing a functional, usable, and accessible resource to serve the future needs of a dynamic, diverse scientific community requires careful consideration of the potential topics and scope of future studies. This design process benefits from the engagement of scientists outside the original team. In the CRN, we observed that the process of thinking about future research questions primarily involved members of the CRN themselves, which hindered the identification of innovative future uses and narrowed the breadth of data incorporated into the VDW. This then limited the usefulness of the data resource for projects led by scientists outside the CRN.

Whereas specific research projects are usually conducted by a small group of scientists with shared interests, creating a maximally useful resource requires the involvement of a broad array of stakeholders. This includes scientists from diverse domains and interests who may use the resource in the future; nonscientists who may have responsibility for data, biospecimens, or other components of a potential research resource; funding agencies that may wish to encourage projects using or extending the resource; and community members who could potentially benefit from the research enabled by the resource. We recommend an experienced, neutral facilitator to lead this process, someone with the skills to elicit and draw together disparate perspectives and priorities. We further encourage funding agencies to provide support for all stakeholders involved in these efforts and to fund in-person meetings to facilitate productive interactions and timely progress.

Complexity versus ease of use

Research infrastructures that grow out of networked science bring together data, biospecimens, research staff, and other components from multiple sources. The heterogeneous nature of these sources is often one of the most valuable aspects of a research resource, yet the resulting complexity can be a substantial barrier for scientists posing new questions. Seeking to make the data as useful as possible, CRN scientists and staff pooled an enormous volume of data from numerous administrative and clinical systems. Each specific research project then had to convert data collected for nonresearch purposes into research-ready data, a months-long process in which staff at each site documented the exact meaning of each data element and how it was coded, identified and resolved outliers and missing data, and assessed temporal changes in data availability and quality. The resulting expense and delays hindered use of data from multiple CRN sites and thus reduced the benefit of the research resource.

Although it is difficult to make data, biospecimen, networks, and other research resources simple, there are ways to make these resources, including the accompanying documentation and processes, less complex and more accessible to a broader audience. While collaborating with network scientists on using the data can help, it can also lead to additional challenges, detailed in the next section. The research scientists involved in creating research resources can benefit from the knowledge and skills of other fields in adapting a resource for expanded use. For example, in creating data resources, clinical informaticists could provide expertise during the data conversion process and information scientists could develop tools to support data curation and documentation. Experts in the field of human-centered design can build user-friendly interfaces and tools designed to make exploring and accessing resources less burdensome. We suggest that domain scientists and funders identify and support experts who can facilitate the creation of useful and usable research resources.

Individual professional goals versus research resource priorities

While scientific networks are often treated as entities in and of themselves, they are composed of individuals with their own

Published OnlineFirst August 28, 2019; DOI: 10.1158/1055-9965.EPI-19-0043

professional goals, which may conflict with the priorities of developing a research resource. In the example of the CRN, we observed that the network goal of increasing data access for external scientists had little direct benefit for individual CRN scientists charged with supporting others seeking to use the data. Similarly, whereas external scientists, funders, and other stakeholders may wish to prioritize rapid conversion of a project-specific infrastructure into a research resource, project scientists may want to share data, biospecimens, and access to participants only after completing their work. These two examples reflect a fundamental tension between providing a service that meets user needs with the desire to do scientifically engaging work that advances a field of inquiry and provides an opportunity for career growth (i.e., promotion and tenure). Both viewpoints have merits yet are intrinsically in conflict.

The rise of team science has begun to shift universities and other research organizations toward promotion and other recognition activities that consider contributions beyond authorship order and strongly encourage scientists to list the creation of research resources on their biosketches and curriculum vitae. However, adoption has been slow, and peer review continues to value more traditional accomplishments. We believe the scientific community, including funding agencies, must do more to move beyond a "class system," in which the ultimate accomplishment is grants and publications in a focused topical area, to an approach that similarly values the contribution of scientists with the desire and skills to create research resources that influence science more broadly. A critical first step is funding support for research resource creation, which will compensate scientists for this work and relieve pressure to support themselves with project-specific work. Another possibility is awarding formal, nonauthorship credit to scientists responsible for creating a research resource on which a publication is based, although this might inadvertently reinforce the devaluation of the contribution of scientists who developed the resource. Funding agencies could encourage, or even require, that grant applications include information about plans for making proposed research resources broadly available at the end of the project, potentially also requiring sustainability plans and usage data in grantee progress reports. Creation of an online database of research resources could make identification of resources easier and provide recognition for the scientists who created those resources but could also contribute to a tiered system of valuing scientific contributions. We assert that adequate reward and recognition is both the most important and most difficult challenge in incentivizing scientists to engage in the creation of research resources.

Conclusions

The NCI-funded CRN faced several challenges as it sought to transform itself from a project-specific infrastructure to a research resource available to the broader scientific community. In the specific case of data, creation of widely useful resources was hampered by limited input from scientists outside the Network; unavoidably complicated data of variable quality within and across sites, and over time; and prioritization of CRN scientist-driven work over resource creation activities. These challenges are consistent with what we have seen in other networks and can be categorized as three challenges inherent in the development of research resources: (i) project-based infrastructure versus a research resource for a broader research community, (ii) complexity versus ease of use, and (iii) individual professional goals versus research resource priorities.

The potential solutions to these challenges are not straightforward. Although acknowledging that they exist can help ease the process of converting project-based infrastructures to research resources, implementing real solutions will require expanded engagement of scientists and other stakeholders, involvement of experts with cross-cutting skills such as informatics, and increased recognition of research resource creation as a legitimate and meaningful contribution to science. Peer reviewers and funders must consider resource creation a natural extension of research projects that benefits from thoughtful planning, inclusion in study timelines, and adequate funding. Institutions and human subjects could further encourage research resource creation by recognizing that their contributions can be maximized through appropriate use of their information to answer questions beyond those of the original study in which they participated. Finally, the challenges described here are not unique to NCI-funded networks or cancer-related projects, so we encourage a broader conversation across the scientific community to develop solutions.

Opportunities to convert project-based infrastructure to research resources for the broader scientific community will continue to increase as the prevalence of large projects and research networks continues to grow. For example, numerous NCI-funded Cancer Moonshot[SM] initiatives involve investment in new networks and the infrastructure to support them (https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative/funding). It is probable that these networks will generate massive amounts of data, biospecimens, and institutional collaborations that could be used for decades to come in ways we have yet to imagine. Ongoing limitations in research funding will likely increase the need to create broadly usable research resources rather than build new infrastructure for specific projects. Overcoming the challenges in converting project-specific infrastructures to research resources will ultimately require scientists and funders to shift from viewing projects as self-contained and time-delimited activities to recognizing that many projects have potential uses far beyond those originally envisioned.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Disclaimer

The opinions expressed here are those of the authors and do not represent the official position of the NCI, NIH, Department of Health and Human Services, or any other federal agency.

Authors' Contributions

Conception and design: B. Rolland, A.M. Geiger
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): B. Rolland
Writing, review, and/or revision of the manuscript: B. Rolland, A.M. Geiger

Acknowledgments

Support for this work came in part from the University of Wisconsin Carbone Cancer Center Support Grant P30-CA014520 (to B. Rolland), NIH, NCI.

Received January 9, 2019; revised April 5, 2019; accepted July 30, 2019; published first August 28, 2019.

Published OnlineFirst August 28, 2019; DOI: 10.1158/1055-9965.EPI-19-0043

References

1. Chubak J, Ziebell R, Greenlee RT, Honda S, Hornbrook MC, Epstein M, et al. The Cancer Research Network: a platform for epidemiologic and health services research on cancer prevention, care, and outcomes in large, stable populations. Cancer Causes Control 2016;27:1315–23.
2. Wagner EH, Greene SM, Hart G, Field TS, Fletcher S, Geiger AM, et al. Building a research consortium of large health systems: the Cancer Research Network. J Natl Cancer Inst Monogr 2005;(35):3–11. Available from: https://academic.oup.com/jncimono/article/2005/35/3/921886.
3. Ross TR, Ng D, Brown JS, Pardee R, Hornbrook MC, Hart G, et al. The HMO Research Network Virtual Data Warehouse: a public data model to support collaboration. EGEMS 2014;2:1049.
4. Bietz MJ, Lee CP. Collaboration in metagenomics: sequence databases and the organization of scientific work. In: Wagner I, Tellioğlu H, Balka E, Simone C, Ciolfi L, editors. ECSW 2009: Proceedings of the 11th European Conference on Computer Supported Cooperative Work; 2009 Sep 11–17; Vienna, Austria. London (UK): Springer; 2009. p. 243–62.

Cancer Epidemiology, Biomarkers & Prevention

AACR American Association for Cancer Research

Addressing Challenges in Converting Grant-Funded Infrastructures to Broadly Used Research Resources

Betsy Rolland and Ann M. Geiger

Cancer Epidemiol Biomarkers Prev 2019;28:1559-1562. Published OnlineFirst August 28, 2019.