Building Global Genomics Initiatives and Enabling Data Sharing: Insights from Multiple Case Studies

Federica Fusi,[1] Daniele Manzella,[1] Selim Louafi,[2] and Eric Welch[1]

Abstract

This genomics global governance research study presents the dynamics and the evolving nature of salient challenges that global genomics initiatives encounter in designing new models for data management, exchange, and collaboration across disciplines, sectors, and countries. Using a multiple case study approach, we assessed and compared organizational responses across diverse genomics initiatives. The richness of a comparative qualitative analysis clearly shows the complexity addressed by genomics initiatives and, importantly, expands current studies by moving beyond an open versus property regime dichotomy. Although we identify some common themes, fundamental differences emerge in the way genomics initiatives set goals, manage heterogeneity, define resources, devise governance, and enable data sharing. Such differences demonstrate the ongoing processes of adapting governance structures, management processes, and organizational design solutions that are implemented in response to different social, technical, and policy environments. We find that genomics initiatives largely benefit from and are shaped by the engagement with large communities of scientists to rethink and design shared rules and guidelines for data exchange and use. Our study provides direct guidance to future global genomics initiatives, but it also offers a benchmark for research in the omics field broadly, both in terms of design and methodological approaches to understand the emerging forms of scientific governance and innovation ecosystems.

Keywords: global genomics initiatives, organizational design, data sharing, case study, global governance innovation

Introduction

"But, you know, you touched on the cultural differences and things like that, […] we can't even get scientists to share data with each other. Let alone put their data into some sort of black box, where they're afraid that their fellow wheat breeder who sits next door to them might actually see what they're doing. So, you know, the challenges there are huge."
Manager at a genomics initiative
Personal communication 2015

Gᴇɴᴏᴍɪᴄꜱ ʀᴇꜱᴇᴀʀᴄʜ ᴘʀᴀᴄᴛɪᴄᴇꜱ have diversified and expanded worldwide over the past decade. This was, in part, due to new bioinformatics tools, distributed online computing across institutional, geographic, and temporal barriers and reduced costs and time for producing, analyzing, and managing genomic data (Balasopoulou et al., 2017; Edwards et al., 2011; Foster and Sharp, 2007; Goff et al., 2011; Nielsen et al., 2009). Yet scientists and researchers increasingly perceive that improvements in information technology (IT) infrastructures and tools are a necessary but not sufficient condition to conduct genomics research. Innovation in global governance of science, technology, and innovation in society is also required.

In this study, we aim at developing a deeper understanding of the current governance systems used in global genomics initiatives. Genomics initiatives are team-led communities designed to enhance access to data for the pursuit of individual research goals and the development of knowledge and innovation that benefit the collectivity. Studies show that building and maintaining global genomics initiatives is significantly challenging because of their open and voluntary nature and despite significant funding from the public and nonprofit sectors many initiatives fail to achieve their goals (Chokshi et al., 2006; Including et al., 2013; Kane and Ransbotham, 2016; Kosseim et al., 2014; Parsons et al., 2011). For this study, we selected six genomics initiatives

[1]Center for Science, Technology and Environmental Policy Studies, School of Public Affairs, Arizona State University, Phoenix, Arizona.
[2]French Agricultural Research Centre for International Development (CIRAD), UMR AGAP, Montepellier, France, and AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France.

in both human health and agriculture sectors as case studies and conducted in-depth interviews with executive or high level managers, founders, and staff for each initiative.

Our comparative analysis identified key challenges that consistently emerge across all seven initiatives and likely affect their successful functioning and long-term sustainability. For each challenge, we identified design solutions adopted by the initiatives and, thus, show that some governance design solutions are common across initiatives, while others are tailored to the particular social, institutional, and scientific context.

Importantly, the present study goes beyond traditional analyses and static approaches that contrast public versus private funding of genomics data or proprietary versus open access approaches to analyze a multiplicity of governance design solutions that are currently adopted by genomics initiatives. Our results show a more complex and nuanced picture of the challenges and trade-offs faced by those global genomics communities. In a broad sense, the study contributes to a more detailed understanding of the processes of reflexive governance and midstream modulation of research and development (R&D) activities in large-scale biology, developed previously in the social sciences literature to describe the integration of scientific, policy, and social considerations in program design (Balmer et al., 2015; de Vries, 2004; Fisher et al., 2006; Vermeulen et al., 2013).

Specifically, we ask: What common challenges are faced by genomics initiatives? Which designs do management teams establish to address those challenges and reach their goals (i.e., establish functionality, build and maintain a community, and encourage data sharing)? Our analysis identifies key challenges that consistently emerge across all seven initiatives and could potentially hinder the establishment of genomics initiatives. For each challenge, we discuss solutions adopted by the initiatives. In some cases, solutions are similar across initiatives showing the existence of common patterns; in other cases, solutions are specific to the organizational and environmental context, suggesting that a one-size-fits-all approach might not work for the design of effective genomics initiatives. We conclude with a discussion of future research and policy implications.

Materials and Methods

This study adopted a multiple case study approach and a comparative methodology to investigate the contexts, decisions, and structures of organizations such as genomics initiatives (Eisenhardt, 1989; Eisenhardt and Graebner, 2007; Schramm, 1971; Yin, 2011). Case studies are most appropriate when the topic of research is defined broadly, contextual characteristics are of high importance, and the research aims to investigate a contemporary phenomenon for which existing evidence and theory are limited (Stake, 1995; Yin, 2011). We rely on interviews cross-checked with written material such as reports or documents as data for addressing our questions. We apply a comparative approach to distill challenges and solutions.

We applied a purposeful sampling strategy to select cases. Case selection not only ensured consistency across key dimensions of interest but also captured reasonable variation across cases to understand how initiatives respond to diverse conditions (Patton, 1990; Stake, 1995). We selected initiatives that:

- explicitly aim to promote data, information, and knowledge sharing in human, plant, or animal genomic research, but do not focus only on a single gene disorder or on narrowly focused genetics topics;
- have multiple goals, such as research, innovation, community building, and service provision (i.e., open repositories that provide only storage space were excluded);
- represent a range of disciplines and subdisciplines of science;
- involve actors from public, nonprofit, and private sector;
- promote global resource exchange across scientists and organizations in different world regions, including developing countries.

After an extensive online search that included input from field experts, such as policy-makers and researchers working on genomics initiatives, and reviews of previous studies, we compiled a list of 26 potential cases that matched our selection criteria. We evaluated each case against the five dimensions of interest and reduced the list to 11 cases. For each, we then scheduled a preliminary semistructured Skype interview with the Executive Director or Project Manager of the initiative. From the interviews, we collected additional information about the initiative goals and activities, participants, and management structure. Based on such additional information, we excluded initiatives that were too narrowly focused, involved a small homogeneous community, or offered very limited additional variation to the final sample. The selected case studies, all of which are related to genomics or high-throughput genomic data are presented in Box 1.

Data collection

We designed an interview protocol to collect information on the organization history and context; data access, use, and exchange; terms of contribution and use; governance; management; outputs and constraints; and gaps and challenges. We refined the protocol after the preliminary interviews. Both preliminary interviews and the final protocol were approved by the Institutional Review Board at Arizona State University. A consent form was provided to the invited interviewees before the interview, and oral consent was obtained by the interviewer before starting the interview. Participation was voluntary and interviewees could decide not to respond to a particular question, stop recording, or end the interview at any point. Participants' identity is kept confidential, and transcripts were anonymized for analysis.

We conducted semistructured interviews with multiple members within each initiative to collect data on diverse perspectives. For most cases, interviewees included individuals in a leadership position such as the primary investigator, top management, and officers in charge of scientific and technical support. We also interviewed individuals in charge of community development, memberships, and communication (including those responsible for developing country interactions) and those involved with commercialization or private partners. When relevant, we conducted interviews with key members of the initiative, particularly the private sector partners.

Each interview lasted ~60 min. We interviewed a total of 26 individuals, between 4 to 6 individuals for each organization. Given the small size of management teams, this number of interviews was sufficient to reach saturation of information (Herbert and Irene, 2004). In two cases, we

Box 1. Summary of Analyzed Initiatives

Initiative A is a service provider, which aims at supporting data-intensive life science research by providing a highly flexible and customizable platform with a user-friendly interface and tools for the management, analysis, sharing, visualization, and cloud storage of large amount of genetic data. Through the platform, scientists can share large-scale databases within their research teams or make them publicly available to other scientists. This latter option is completely voluntary. As periphery activities, the initiative also offers information technology support to users; acts as a broker to help establishing collaborations among members; and occasionally partners with members to develop grant proposals.

Initiative B aims at facilitating joint precompetitive research activities in healthcare among for-profit companies. Research goals are anonymously established by members of the initiatives, and all research activities are carried on by a partner university. Research activities generally focus on common issues that are too expensive to be undertaken by individual actors. All research results are made publicly available after 2 years with no restrictions on their use. Openness is a mandatory rule for all members. Members anonymously contribute with data and other information in order to facilitate joint research activities.

Initiative C targets scientists and breeders in developed and developing countries to encourage the deployment and adoption of new technologies that enable and facilitate traditional and molecular breeding. It offers a flexible data management platform along with local and online training to scientists and breeders from developing countries. The initiative is organized in a hub-and-spoke structure, with several regional communities all around the world.

Initiative D is an online distributed platform that facilitates access to computational capacity to research facilities, predominantly around the United States. Research institutes looking for capacity are matched with slots of computational capacity provided by other research institutes. Providers of computational capacity can establish at which conditions users can access and utilize their resources. Requests are matched by an online algorithm. Research institutes can either request or provide computational capacity or both.

Initiative E aims at building a community engaged in designing, disseminating, and adopting common standards and principles for data sharing in human genetics and genomics. The initiative involves several actors from different sectors and world regions. Members work within organized working groups, each of which focuses on a specific topic. Solutions developed in the working groups are developed and tested as part of small demonstration projects to demonstrate the value of a shared framework for data sharing.

Initiative F is building an open access repository of rice genetics data. It combines several public datasets while supporting and conducting gene sequence activities. It also facilitates collaboration and communication across members of its reference community. While access to the platform and common data are free, the initiative is evaluating fee-based options for for-profit private companies, which want to get early access to new data.

encountered difficulties in conducting interviews. In one case, members were reticent to talk because of competition concerns. We partially mitigated such concerns by conducting anonymous phone interviews. In another case, we faced resistance from the leadership concerned about burdening managers, who were mostly volunteers. Accordingly, we limited our interviews and supplemented our data collection with documents provided by the organization.

Analysis

Triangulation across diverse sources of information is key to increase the internal validity of findings and the accuracy of conclusions (Yin, 2011). We integrated interview data with documents retrieved from the initiatives' website or provided by the interviewees. Documents include meeting summaries, constitutive agreements, letters of agreement signed by members, policies, guidelines, and rules. Documents were used to confirm events and facts reported by interviewees and to complete our understanding of the structure and evolution of the initiatives.

A case study protocol was used to systematically analyze the data collected by theme and provide a general understanding of each case under investigation (Yin, 2011). The case study protocol was revised by all team members that conducted the interviews (two to three researchers for each case). The research team then conducted a pattern matching analysis identifying common challenges and solutions within each case and communalities across cases. Evidence was extracted by checking and continuously validating text in interview transcripts. Identified themes and related evidence were discussed at the team level until agreement was reached regarding common challenges and solutions.

Results

Five Challenges for genomics initiative

> "It is an enormous job that requires very innovative approaches that have probably not been used before. Because we are dealing with very unique challenges to get this job done. And the starting point for a lot of the organizations that we are dealing with varies very widely."
> Manager at a genomics initiative
> Personal communication 2015

Genomics initiatives are generally characterized by an open structure, in which scientists and other organizations are free to participate, share, and use resources, while pursuing their own goals. Emphasis is put on open access approaches as a way to counter proprietary institutions and norms that are deemed to limit scientific progress (Kosseim et al., 2014). Actors are autonomous, but interdependent with one another because each individual contribution increases the overall outcome for the community. However, since participation is voluntary, actors might drop out or free-ride, using but not providing common resources. Such behavior negatively affects the community as the continuous production of new content is necessary to maintain the community over time (Kane and Ransbotham, 2016).

Researchers have investigated how genomics initiatives manage their community, which rules they establish for guaranteeing use and contribution of resources, and how they

design a common infrastructure and address conflicts among members (Chokshi et al., 2006; Including et al., 2013; Kosseim et al., 2014; Parsons et al., 2011). We identified five common challenges across the six initiatives as follows: (1) setting goals, (2) managing heterogeneity, (3) defining resources, (4) devising governance structure, and (5) enabling data sharing. We describe each challenge in this study, and the Table 1 highlights each of them.

Setting goals

We found that genomics initiatives struggle in developing a clear and concrete mission as they are challenged to integrate different goals based on the priorities of the diverse communities they aim to attract.

The initial vision most often emanates from a prominent scientist or a small group of scientists, who desire to fill gaps not addressed by existing funding systems and organizational structures. The initial vision is often very broad and ambitious because founders are trying to attract diverse and numerous stakeholders who have a wide range of needs, recruit complementary skills across disciplines, and aim to achieve high-level goals, including open data sharing or knowledge coproduction. As a result, the initial vision is often too complex to be implemented, and initiatives at early stages encounter several difficulties in moving from the founders' vision to an actionable plan.

Three types of goals typically match the initial vision. *Research goals* aim to leverage the stock of technological developments to synthesize common knowledge and promote research innovation. *Community building goals* emphasize the iterative aspects of collaboration and promote interactions among community members to coproduce knowledge at multiple scales and in multiple locations. Finally, *service provision goals* tend to avoid interference with existing collaboration structures and norms and, instead, aim to neutrally provide services to support data-intensive research. Tensions emerge in the implementation of these three main goals: prioritizing one goal rather than others shapes resource allocation, community, and activities.

Managing heterogeneity

Integrating a variety of actors with substantial disciplinary, sectoral, and geo-social differences creates tensions and conflicts within the community. We found that heterogeneity is the main challenge that genomics initiatives need to address.

We identify three main types of heterogeneity—disciplinary, sectoral, and geo-social—each of which leads to collective action and coordination problems. Disciplinary heterogeneity can lead to clash of culture and values across individuals in different fields. It is often the case between infrastructure science (hardware), discovery science (breeders and geneticists), and bioinformatics (software). Sector heterogeneity leads to a higher risk of free-riding behaviors among actors because it increases coordination challenges and the need to integrate different sector-based norm values. Finally, heterogeneity of capacity embedded in geo-social diversity raises distributional conflicts related to input allocation and resource redistribution. In our case studies, this is typically manifested in the relationships between developed and developing countries.

In all cases, management staff report difficulties reconciling various types of heterogeneity, especially when heterogeneity is extremely high. High heterogeneity is a challenge because of the conflicting goals and diversity of needs that characterize the interactions among actors involved. When not addressed, heterogeneity hinders the initiative's ability to move forward because expectations of the different participating communities are not fulfilled.

Defining resources

Initiatives combine and generate multiple resources that are difficult to manage and maintain. Initiatives depend on resources voluntarily provided by multiple sources holding different interests, thus presenting the risk of free-riding behavior.

Resources can be of different nature, including scientific resources such as genomic or phenotypic data and biological

TABLE 1. SUMMARY TABLE OF KEY CHALLENGES IN GENOMICS INITIATIVES

Key challenge	Dimensions	Critical issues
Setting goals	Research Community-building Service provision	Balancing goals and interests of the different communities that the initiative is trying to involve while setting achievable and implementable goals.
Managing heterogeneity	Geo-social Disciplinary Sectoral	Promoting trust and reducing cognitive barriers and conflicts across heterogeneous actors and preventing free-riding behaviors that damage resource exchange.
Defining resources	Technical Scientific Human Social Managerial	Developing and maintaining a sustainable flow of resources drawing from voluntary contributions of community members and competitive grants.
Devising governance structures	Hierarchy Representation Competence	Devising a structure that ensures legitimacy and authority to the initiative, while balancing representation, efficiency, and quality of decision-making.
Enabling data sharing	Organizational and technical resources Institutional capacity	Designing an appropriate set of incentives to overcome resistance to data sharing among community members, considering the available organizational and technical resources and the initiative's institutional capacity.

material or technical resources such as software (analysis tools and application programming interfaces [APIs]) and storage space. But there are also human and social resources, such as staff, to help with IT infrastructure and facilitate interaction and deliberation, and relational resources—connection, trust, reciprocity, and support—within the network of individuals that make up the initiative membership.

First, initiatives compete over scarce financial resources, such as public funding, that provide for only some of the potentially needed resources. Second, access to and use of common resources often does not require a matching equivalent provision of resources from members. Contribution by participants remains voluntary. As such, resource flows are uncertain and there is a high risk that members adopt free-riding behaviors: using resources but not contributing to the pool. Genomics initiatives struggle in defining their resources and finding ways to maintain stable flows of needed resources.

Devising governance structures

Governance defines the structure of deliberation and participation processes in genomics initiatives, including definitions of authority and legitimacy. Genomics initiatives struggle to balance the need to represent all different stakeholders, ensure adequate technical competences, and implement efficient processes. Genomics initiatives need to consider trade-offs among authority sources in their governance structure.

Governance describes the arrangement of the committees, groups, and units that guide and direct the initiative, thereby shaping its deliberation and participation processes. It also defines the sources of authority that provide legitimacy to the decisions made by such committees, groups, and units and their role of monitoring and enforcing compliance among members.

Some genomics initiatives leverage hierarchy, which refers to the authority placed in the position of an individual, group, or office within an organization. Hierarchical organizations enable one person or group to make decisions for the entire community. Other initiatives rely on representation- or competence-based authority. Representation concerns the extent to which relevant stakeholders either agree that their interests are represented by others (i.e., by electing representatives) or are directly involved in decision-making. Representation reflects a multiplicity of constituencies, which are typically tied with political and legal uncertainties surrounding the governance of the resources. Competence recognizes knowledge, experience, or professional credentials of some members.

Selecting one source of authority over the other significantly impacts the efficacy and outcomes of the decision-making process. While leverage of representative authority increases legitimacy among community members, it also slows down decision-making compared to hierarchy or competence authority. But decisions based on hierarchy or competence might be less accepted by community members.

Enabling data sharing

Genomics initiatives aim to promote data sharing among community members. Promoting data sharing requires considerable effort, and genomics initiatives often experiment with different models that integrate incentives and rules for encouraging data sharing.

Data sharing is often hindered by several social, cognitive, and institutional barriers, including scientists' concerns about competition and data misuse; differences across fields in standards and metadata; or lack of compatible technical infrastructures, among others. Genomics initiatives need to design instruments and incentives to encourage data sharing among community members. Rules and incentives vary significantly across cases, often contingent upon the resources mobilized by the initiative.

We observe that initiatives evaluate data sharing approaches by considering the organizational and technical resources that they might provide to members and their institutional capacity to enforce common rules that allocate rights to use, access, and share data. Organizational and technical resources include storage space and IT infrastructures to promote accessibility and data exchange in the first place.

Some initiatives also offer APIs and software tools to facilitate interoperability across members' own systems. Others help community members design metadata and standards. Accessibility to such resources provides incentives for members to share and contribute data. Institutional capacity determines whether an initiative adopts common rules for sharing data among members or allows members to set their own rules. The balance between autonomy and common rules is particularly delicate because members might refrain from joining initiatives whose rules do not match with their perspective. At the same time, the lack of rules might hinder the development of trust among members and, as a consequence, data sharing.

Design solutions for genomics initiatives

"I was looking at the incentives [to encourage data sharing] that have been listed in the proposal. […] the reality is that there's only about one or two of those that really work."
Manager at a genomics initiative
Personal communication 2015

We identify several solutions for each challenge. The diversity of solutions highlights the complexity of genomics initiatives and data sharing, both of which are fundamentally embedded in the social and organizational structure of scientific research. It also emphasizes managers' recognition of their specific social, technical, and political environments which significantly shape preferences over certain solutions compared to others. Initiatives define boundaries, goals, resources, and funding in accordance to the context, as well as the feedback from the community. A summary of the design solutions discussed in the next paragraphs is presented in Table 2.

Goal setting

Starting big, going small.   The establishment of genomics initiatives is best characterized as step-wise, evolutionary, and fraught with significant challenges that result in reassessment and change of trajectory. All interviewers consistently report that they needed a few years to transition from an idea to an actual functioning project; all began with a much larger and comprehensive mandate than they have currently. By narrowing the scope and aims of the initiative,

TABLE 2. SUMMARY OF IMPLEMENTED SOLUTIONS

Key challenge	Solutions
Goal setting	Take time (even few years) to move from a broad ambitious mission to a clear concise one.
	Prioritize one goal (research, community-building, and service provision) and design others as support.
	Clearly define initiative boundaries and niche.
	Depending on short- or long-term goals, opt for a demand-pull or supply-push approach.
Heterogeneity	Design homogeneous subcommunities which are more effective in managing collective action.
	Integrate diverse communities over time, once the initiative has established its functioning rules and structure.
	Engage heterogeneous groups in small short-term projects that provide great rewards for participants.
Resources	Develop a sustainable model by combining different sources of resources: public funding and contribution from members.
	Design core resources and minimum rules to protect them over time.
Governance	Leverage on representation- and competence-based authority to guarantee commitment and trust from the community.
	Evaluate the environmental context—that is, heterogeneity and political tensions—when designing decision-making processes.
	Design a reliable management structure: it helps building trust toward the initiative, shaping expectations, and influencing informal norms among members.
Data sharing	Provide adequate to excellent organizational and technical resources: while they do not ensure data sharing, they are a fundamental precondition.
	Design norms that reflect institutional capacity to enforce them, while paying attention to what might or might not attract the community.
	Integrate different incentives that address organizational, technical, or institutional concerns.

the management team was able to effectively design specific activities and move forward the initiative's goals. Furthermore, because these initiatives are highly innovative in a relatively conservative scientific culture, smaller size and scope give them an opportunity to demonstrate value and effectiveness, reduce complexity, and focus on mission.

Prioritizing goals. None of the initiatives prioritize all three goals equally—research, community development, and service provision; instead initiatives prioritize one goal while integrating others as periphery. For instance, when initiatives develop common research goals (high research integration),

they generally do not provide services, focusing instead on brokering services and expertise among members (medium community-building). Likewise, initiatives that invest in capacity development (high community-building) are more willing to provide IT tools and technical standard (medium service provision) to promote collaboration. Finally, initiatives which provide extensive technical infrastructures (high service provision) engage little in community building and provide only limited support for developing practices and methods for existing research projects (low-to-medium research integration). Prioritizing goals is essential to narrowing the initiative's mission and effectively allocating otherwise scarce resources.

Defining boundaries. Most interviewees were highly cognizant of their initiative's boundaries. Genomics initiatives are continuously faced with options for future growth and development, particularly when they need to secure funding and engage with heterogeneous stakeholders—scientists, bioinformatics experts, and funders, among others. To focus on their mission, initiatives seek to clearly distinguish core and periphery activities, interests, resources, and services; in other words, the initiatives clearly define what is and is not required for accomplishing their vision. The management team usually requires significant time and energy to articulate, establish, and maintain their niche, which defines the uniqueness of the initiative and its competitive advantage.

Demand orientation. Initiatives adopted one of two fundamental design patterns—supply-push or demand-pull—depending on the extent to which the needs of community members drive the goals of the initiatives. For a supply-push design, the management team establishes the direction and activities of the initiative without integrating *a priori* the users' perspectives. By contrast, the management team focuses on meeting the needs and interests expressed by the community for a demand-pull design. The two designs are end points of a continuum.

All initiatives show characteristics of both designs, although they predominantly tend toward a demand-pull design. Adopting a demand-pull design requires longer lead time to aggregate and address community needs, especially in highly heterogeneous community, but it facilitates commitment toward common goals and the adoption of a shared technology. Once demand-pull initiatives defined their niche in response to users' needs, they did not engage in new activities unless they clearly identify a demand from their community.

By contrast, supply-push initiatives require greater effort and resources to attract and maintain members and disseminate technologies, but the design can be launched more rapidly as critical decisions are taken by fewer individuals. Supply-push initiatives are more likely to change their niche in response to new goals expressed from the management, but they require greater effort to create and define their niche as they need to attract enough support and interest.

Heterogeneity

Subcommunities. Some initiatives manage heterogeneity by creating homogenous subcommunities, which facilitate

coordination and effective joint action. For instance, some initiatives manage geo-social diversity by creating region-based communities, while others create subcommunities based on participants' interests to manage disciplinary heterogeneity. The benefits of this approach, however, are ambiguous. Subcommunities have advantages of smaller size and greater homogeneity, but coordination problems may arise if members restrict collaboration to their comfort zone and do not engage with other subcommunities. When this occurs, initiatives achieve only a fraction of the cooperative potential offered by the community.

A staged approach. Some initiatives choose to start with a single homogenous community to reduce complexity in the design phase. This is helpful for gathering enthusiasm and energy around the initiative and facilitating the establishment of common rules. After consolidation, the initiative expands the initial group to include more heterogeneous actors. Integration over time is more effective as the initiative has already developed its own governance system—structure, decision-making processes, institutions, and norms—which is accepted by the existing community and does not require significant engagement in new negotiations as new groups join.

Showcasing value through projects. Some initiatives engage heterogenous members on small common projects. Such projects demonstrate the value that can be created by collaborating within the initiative's framework. Joint action among members, even on small activities, leads to a virtuous circle in which successful results increase mutual understanding and trust. Such activity also provides the foundation for the emergence of joint aims at a later stage (Huxham and Beech, 2003). We notice that to be effective, projects have to focus on a narrow mission that can be undertaken with minimal resources and hold the interest of actors with different research goals. Moreover, it is fundamental to select short-term goals, which can be reached within a short period of time but offer high rewards for participants. For example, some initiatives engage members in API development.

Resources

A search for sustainable models. We found that no initiative has a concrete well-defined solution to secure funding stability over time. The resources are either developed through grant funding involving public or nonprofit funders, such as National Science Foundation, Wellcome Trust, and the Gates Foundation, or mobilized through voluntary contributions from the community. But the financial model varies across cases, with some initiatives that are entirely publicly funded to others that rely heavily on the voluntary contribution of member time and resources.

It is worth noting that very few initiatives ask for financial contributions from their community members. The few initiatives that apply a fee-based system differentiate across users, not to undermine the open access nature of the initiative. For instance, one initiative grants developing country scientists free access to resources, while it asks for a fee to scientists and institutions from developed countries. Another initiative charges a low fee to industry members, who in return are allowed early access to precompetitive innovation outputs of the initiative. The design of the funding system is more often a response to the social environment in which those initiatives are embedded.

Focus on core resources. Since resources are costly and funding is hard to maintain, genomics initiatives generally focus on one or two core resources while other contingent resources provide support. Access to core resources is sometimes subject to minimum requirements to protect the common pool from free-riding behaviors—that is, excessive consumption of computational capacity resources or illegal download of data from shared servers. The requirements can be as low as providing an institutional e-mail address before downloading the data.

Governance

Representation- or competence-based authority over hierarchy. Genomics initiatives show a balance of different sources of authority. Hierarchical authority is generally low and of limited importance. Genomics initiatives rely on strong leaders only when they need to give a strong identity to the initiative; consensus-based decision-making becomes too slow and unproductive; or the initiative is undergoing major changes that require rapid decision-making and clear leadership.

By contrast, representation- and competence-based authorities vary substantially across cases, depending substantially on the initiative's goals. Initiatives which lean toward service provision tend to rely on competence-based authority, particularly when the product is inherently technical. Community-building initiatives depend on representation-based authority, as they need to establish legitimacy among community members. Representation generates trust and buy-in by the membership and may benefit long-term sustainability of the initiative. Yet we noticed that representation can also be an obstacle to efficient decision-making. In initiatives where different actors with diverse interests hold authority, decisions are prone to debate, compromise, and synthesis, increasing time and costs of decision-making processes.

For the most part, initiatives try to balance competence and representation by appointing bodies that rely on both authorities. For instance, some initiatives are led by a coordinating manager who is the head of a small team of individuals respected for their scientific expertise and valued as independent thinkers. At the same time, an external advisory board or steering committee is created as representation-based body to maintain contacts with key stakeholders and their perspectives.

Centralized or decentralized decision-making. The design of genomics initiatives includes either a decentralized or centralized approach to decision-making. Genomics initiatives adopt centralized decision-making when they operate in a politicized context involving a wide range of stakeholder interests, when norms are unsettled or evolving, and when the initiative aims to develop a single product or service for heterogeneous actors. Some initiatives temporarily adopt a centralized approach during an early or transitional stage. Under such conditions, a centralized approach can facilitate communication among key stakeholders and provide stability, such that the initiative is able to operate and coordinate

joint actions. A decentralized decision-making approach might undermine the initiative's success as priorities and standards would be unstable and depend on representation and power dynamics.

Decentralized decision-making mostly occurs in contexts where collaboration on research is the primary goal. In this study, decision-makers are highly skilled professionals who make decisions based on professional norms and standards. The decentralized model is also more common when there is high geo-social heterogeneity, which could lead to high communication and information costs in a centralized structure. Finally, decentralization is more likely in cases where the initiative seeks innovative approaches to problem solving that are not well understood by single individuals.

Management as one component of governance. All our initiatives report considerable investments in a reliable and efficient management, which does not necessarily coincide with the leadership of one individual. Interviewees consistently agreed that designating an effective management team is necessary to ensure production of a high-quality product that satisfies community members. Good management also improves the quality and reliability of internal processes that address sensitive issues such as security and intellectual property of data, thereby increasing trust toward the initiative and across members.

The composition and structure of the management team depend on the goals, membership, and resources embedded in the initiative. In most cases, a small team of highly respected individuals with demonstrated competencies in their field constitutes the management team. Management teams are particularly effective when they have experienced members and have sufficient resources such that they can demonstrate the value of sharing and collaborating from initial stages.

Finally, management shapes the actual functions that the initiatives implement. It is the performance and practical application of those functions that shape structures, expectations, and norms among members over time. Given their innovative nature, the governance of genomics initiatives is most often created through managerial practice rather than by formal design. Over time, once the initiative has clear goals, explicit guidelines and rules may become more important.

Data sharing

Autonomy of sharing data. Few of the initiatives ask members to share their data as part of membership requirements. A mandatory requirement for sharing data might prevent members from joining, thereby limiting the ability of the initiative to build a solid community, which is the primary goal for most initiatives. Interviewed managers share the belief that data sharing will happen as the community consolidates and grows over time.

Some genomics initiatives set guidelines on how to share data. Guidelines are not enforced but provide a common reference among members on appropriate behaviors for sharing data. Guidelines might be specific to organization-related activities or might provide broad suggestions on data sharing. Other initiatives require members to make data publicly available when such data are produced as part of the initiative activities, that is, when members collaborate on joint projects. Initiatives are more likely to develop common rules for data sharing when they have resources (i.e., technical processes or dedicated teams) to provide incentives, ensure compliance, and enable monitoring among members.

Organizational and technical resources. Providing access to technical support is crucial to create incentives for members to share data. Most initiatives provide open unlimited storage for users who deposit data and openly share them. Moreover, most users need technical support to upload data or develop appropriate metadata and data sharing standards. In some cases, technical teams even help users with skills necessary to analyze or visualize data. Providing such technical services is usually an incentive for data sharing: members are more willing to share data if they receive access to complementary skills, technical expertise, and social relationships.

Table 3 summarizes the incentives for data sharing utilized by our cases. We provide a brief description of their effectiveness and conditions under which they may be applied according to evidence reported across interviewees.

Discussion

This study presents a fresh form of inquiry into science management and governance. Genomics research is a complex process that requires collaboration and sharing to integrate several inputs to research, including data, scientific, and informatics skills, computational and storage capacity, and biological materials. While IT improvements reduce technical challenges of long-distance collaboration and sharing, they do not remove institutional, social, or behavioral barriers that prevent them (Eisenberg, 2006). Among others, competition for discovery and publication, conflicting interests, capacity gaps, lack of standards, and common norms still discourage scientific collaboration and open sharing of resources (Campbell and Bendavid, 2003; Hedstrom et al., 2008; Piwowar and Chapman, 2008; Shibayama et al., 2012).

Public agencies, foundations, and other private entities are increasingly supporting global genomics initiatives that aim to promote resource sharing, especially data, and collaboration for genomics research by simultaneously developing technical infrastructure and addressing social, institutional, and organizational barriers. The recurring consideration is that data sharing barriers cannot be addressed in isolation or within small research groups (Including et al., 2013). Open access IT infrastructures, centralized services, storage space, and applications are only one step toward sharing resources among geographically dispersed organizations and scientists and supporting both smaller research program and big science (Alberts, 2012). Genomics research benefits from the engagement with large communities of scientists to rethink and design shared rules and guidelines, formal and informal rules of data exchange and use, metadata, and common standards.

Genomics initiatives—and other similar projects—represent an important research setting to understand fluid midstream approaches to technological development in science, where decisions are less constrained by policy agendas or linear approaches to scientific and technological change (Fisher et al., 2006), but are collectively discussed and modulated by scientists. Yet policy analysis of genomics initiatives is commonly associated with tensions between proprietary and open access approaches. All the other dimensions, such as the

TABLE 3. INCENTIVES FOR DATA SHARING IN GENOMICS INITIATIVES

Incentives	Description
Storage space and IT infrastructure	By providing storage space and IT infrastructures, genomics initiatives support data sharing as users and contributors can access IT services, data analysis tools, and space to upload data.
Facilitating units/ organizations	Facilitating units are small dedicated teams that support data sharing by bridging actors and creating the social structure needed for enabling sharing. Demonstration projects can be facilitating units. It is important that facilitating units define and promote equitable sharing among members.
Data pooling	Members are equally required to contribute to the pool to access it. This instrument works best when actors derive benefits from accessing the pool and have equal capacity to contribute. Data pooling can be combined with other instruments, such as facilitating units or demonstration projects. It might create problems in highly heterogeneous initiatives.
Access to technical expertise	Given technical skills required for genomics research, the provision of technical support to manage and analyze data or conduct other activities is a strong incentive for sharing data. Many initiatives assist members with technical expertise under the condition that data and analyses will be publicly available after an embargo period.
APIs, software tools: interoperability and accessibility	APIs and other software tools promise to solve issues of integration and accessibility of datasets. Most of the analyzed initiatives aim to develop tools that enable easy integration and access of members' own datasets.
Data quality	High-quality data are an incentive for data use and contribution. Initiatives that offer additional analyses on data or ensure the quality of data provide an added-value service to users. Members have incentives to contribute to gain access to additional data analyses or high quality data. Yet, initiatives generally discharge responsibility on the quality of data hosted in their repository.
Access to premium resources/ reduced membership fee	Access to additional resources or a reduced membership fee might act as incentives to share data, especially when the community is established and members are aware of the membership value. Several interviewees refer to the opportunity to offer a reduced-price access to resource in exchange of data contribution. Nevertheless, none of them has yet adopted this approach.
Embargo to first publication or innovation appropriation (early access to data)	An embargo period on data allows contributors to capitalize their investment in producing or collecting data—that is, by publishing or disclosing their invention. Initiatives generally recognize an embargo period on both data that members have autonomously produced or data jointly produced in common projects.
Prepublication/controlled access	Prepublication or controlled access allows researchers to access data before they are published. Those mechanisms speed up scientific discovery without affecting first publication rights of data owners. In prepublication and controlled access the initiatives act as third party and monitors compliance with users and owners' rights among members.
Access to or blockage of funding	Access to funding is a strong motivation for sharing data. NSF policies already require scientists to make data freely available online. Some initiatives implement a similar mechanism, providing part of funding only after members present a data sharing plan. This instrument is effective only if combined with other resources–repository, standard, data quality checks, monitoring, or sanctions.
Development of common standards	Standard development facilitates sharing by increasing trust among community members. It also creates conditions for users to access data and use them in further research. Interviewees suggest that this instrument is more effective when it involves relevant communities in standard design.

API, application programming interface; IT, information technology; NSF, National Science Foundation.

five that we have presented in this article, are often analyzed through the lenses of this dichotomy.

This study offers both a more complex picture of the various challenges and trade-offs that influence the formation and functioning of genomics initiatives, and a recognition of the modulating role that these initiatives play as they seek to balance different intentions, values, interests, influences, and goals (Fisher et al., 2006). Because genomics initiatives facilitate access to common resources such as data storage, manipulation, sharing, and scientific collaboration, they represent a new organizational form that occupies the nexus between small and big science (Alberts, 2012), between

multiple heterogeneous stakeholders, including policy-makers and the broader public (Fisher et al., 2006), and between different visions of future scientific and technological trajectories (Williams, 2006).

From our analysis, it clearly emerges that there is no simple or uniform way to address the challenges that genomics initiatives face generally and especially in their starting-up phase. Given the complexity, interlinked, and context-specific nature of genomics initiatives, it is not appropriate to recommend prescriptive designs or liner approaches. Rather, each initiative faces tensions that exist between alternative solutions and adopts modes of action that are appropriate to

its own context. The richness of a qualitative comparative analysis lies in providing several options that can be adopted according to contextual circumstances.

The study identifies challenges and solutions that can be used as critical touchpoints for socio-technical integration in other genomics initiatives and collaborative science programs (Fisher et al., 2015). For example, solutions to the challenge of goal setting concern primarily initiative scope and assumptions about community needs. Adoption of a broad set of goals and a demand-pull orientation may set high expectations that cannot be met by a leader and a small staff.

Nevertheless, prioritization of goals could involve a range of stakeholders to test the demand and develop socio-technical solutions that meet the support of the community. Similarly, enabling data sharing may require a phased approach in which key users first begin to trust the initiative to securely store and appropriately manage their data, before considering options for sharing with other members of the community. While not comprehensive, the challenges and contingent solutions explored in this study provide scientists with socio-technical insights that may be considered during design and implementation.

The five areas are also interlinked in multiple ways, and trade-offs exist across them. For instance, decisions about core services and activities in the goal setting phase must consider complimentary capacities that are needed to provide contingent services. In this sense, resources form the basis for existence around which the initiatives are organized, including their scope, niche, and complexity. By defining boundaries and setting priorities, initiatives shape their potential membership and the degree of heterogeneity across members, thus determining the need for strategies for managing heterogeneity and devising governance structures that include representation-based authority, decentralized decision-making, and strong management. Initiatives or projects with common rules on sharing entail technical and organizational resources, although with different levels of intensity.

We also acknowledge several limitations of the study. First, our sample is limited to six initiatives. Adopting a purposeful sampling strategy allowed us to highlight some key dimensions of interests but we have not included all possible variations that we could encounter across genomics initiatives. For instance, most of our cases are primarily focused on U.S. scientists and research institutes. Second, by dividing the challenges and solutions into five key areas, we have artificially simplified the complexity of genomics initiatives to deeply focus on specific issues one at a time. While our study still shows that there is great variation in the design and organizations of genomics initiatives, readers should be cognizant that, within such variation, trade-offs and balances need to be struck within the five areas we have selected.

Conclusions

The context-specific nature of genomics initiatives and the diversity of responses associated to the different challenges and trade-offs suggest the importance of trajectory and performance feedback, as additional dimensions that deserve consideration by managers and further research by scholars. For example, mechanisms to monitor progress and ensure adequate corrective actions (e.g., whether further instruments

should be designed to engage with other relevant communities; whether goals and activities should be revised) could smoothen the integration of the five areas that we have explored and allow initiatives to continuously adjust over time. Further research on a reflexive or modulating approach to governance of genomics initiatives should acknowledge the idiosyncrasy of genomics initiatives and strengthen the analysis of the dynamic relationships among individual behavior, institutions, and policies.

Policy-makers should also take comfort in the level of sophistication of initiative design. Evidence shows that genomic initiative leaders are fundamentally able to work through complex policy and social environments to deliver design models that address simultaneous challenges, including ethical concerns and equity issues.

For instance, genomics initiatives pay significant attention to the engagement of heterogeneous stakeholders, representation, and to accessibility issues across world regions and in the face of ongoing biases in funding toward big science projects. Exogenous approaches to governance that depend on linear or anticipatory approaches to scientific and technological forecasting may be less effective than more endogenous, engaged design processes (Fisher, 2005). That said, modulating or reflexive governance does not obviate the need to ensure input and awareness rising from the broader public (Sarewitz, 2015). While genomics initiatives propose a midstream approach to governing and managing R&D, the link between scientist-led communities and the broad public deserves further exploration to ensure the inclusion of diverse cultural and social values into science development (Editorial, 2015; Sarewitz, 2015).

Acknowledgments

Author Disclosure Statement

The authors declare that no conflicting financial interests exist.

References

Alberts B. (2012). The end of ''small science''? Science 337, 1583.

Balmer AS, Calvert J, Marris C, et al. (2015). Taking roles in interdisciplinary collaborations: Reflections on working in post-ELSI spaces in the UK synthetic biology community. Sci Technol Stud 28, 3–25.

Balasopoulou A, Mooy FM, Baker DJ, et al. (2017). Advancing global precision medicine: An overview of genomic testing and counseling services in Malaysia. OMICS 21, 733–740.

Campbell EG, and Bendavid E. (2003). Data-sharing and data-withholding in genetics and the life sciences: Results of a national survey of technology transfer officers. J Health Care Law Policy 6, 241–255.

Chokshi DA, Parker M, and Kwiatkowski DP. (2006). Data sharing and intellectual property in a genomic epidemiology network: Policies for large-scale research collaboration. Bull World Health Organ 84, 382–387.

De Vries R. (2004). How can we help? From 'sociology in' to 'sociology of' bioethics. J Law Med Ethics 32, 279–292.

Editorial [No authors listed]. (2015). After Asilomar. Nature 526, 293–294.

Edwards PN, Mayernik MS, Batcheller AL, Bowker GC, and Borgman CL. (2011). Science friction: Data, metadata, and collaboration. Soc Stud Sci 41, 667–690.

Eisenberg RS. (2006). Patents and data-sharing in public science. Ind Corp Change 15, 1013–1031.

Eisenhardt KM, and Graebner ME. (2007). Theory building from cases: Opportunities and challenges. Acad Manage J 50, 25–32.

Eisenhardt KM. (1989). Agency theory: An assessment and review. Acad Manage Rev 14, 57–74.

Fisher E, Mahajan RL, and Mitcham C. (2006). Midstream modulation of technology: Governance from within. Bull Sci Technol Soc 26, 485–496.

Fisher E, O'Rourke M, Evans R, Kennedy EB, Gorman ME, and Seager TP. (2015). Mapping the integrative field: Taking stock of socio-technical collaborations. J Respons Innov 2, 39–61.

Fisher E. (2005). Lessons learned from the Ethical, Legal and Social Implications program (ELSI): Planning societal implications research for the National Nanotechnology Program. Technol Soc 27, 321–328.

Foster MW, and Sharp RR. (2007). Share and share alike: Deciding how to distribute the scientific and social benefits of genomic data. Nat Rev Genet 8, 633–639.

Goff SA, Vaughn M, McKay S, et al. (2011). The iPlant Collaborative: Cyberinfrastructure for Plant Biology. Front Plant Sci 2, 34.

Hedstrom M, Niu J, and Marz K. (2008). Incentives for data producers to create ''archive-ready'' data: Implications for archives and records management. Proceedings from the Society of American Archivists Research Forum, San Francisco, CA.

Herbert JR, and Irene SR. (2004). Qualitative Interviewing: The Art of Hearing Data, 3rd ed. Sage Publications, Thousand Oaks, CA.

Huxham C, and Beech N. (2003). Contrary prescriptions. Recognizing good practice tensions in management. Organization Studies 24, 69–93.

Including NCDS, Perrino T, Howe G, et al. (2013). Advancing science through collaborative data sharing and synthesis. Perspect Psychol Sci 8, 433–444.

Kane GC, and Ransbotham S. (2016). Content as community regulator: The recursive relationship between consumption and contribution in open collaboration communities. Organ Sci 27, 1258–1274.

Kosseim P, Dove ES, Baggaley C, et al. (2014). Building a data sharing model for global genomic research. Genome Biol 15, 430.

Nielsen CB, Cantor M, Dubchak I, Gordon D, and Wang T. (2009). Visualizing genomes: Techniques and challenges. Nat Methods 7, S5–S15.

Parsons MA, Godøy Ø, LeDrew E, et al. (2011). A conceptual framework for managing very diverse data for complex, interdisciplinary science. J Inf Sci 37, 555–569.

Patton M. (1990). Qualitative Evaluation and Research Methods. Sage Publications, Beverly Hills, CA.

Piwowar HA, and Chapman WW. (2008). A review of journal policies for sharing research data. AHIA Annual Symposium Proceedings, 596–600.

Sarewitz D. (2015). CRISPR: Science can't solve it. Nature 522, 413–414.

Schramm W. (1971). Notes on Case Studies of Instructional Media Projects. Washington, DC: Educational Resources Information Center.

Shibayama S, Walsh JP, and Baba Y. (2012). Academic entrepreneurship and exchange of scientific resources: Material transfer in life and materials sciences in Japanese Universities. Am Sociol Rev 77, 804–830.

Stake RE. (1995). The Art of Case Study Research. Sage Publications, Thousand Oaks, CA.

Vermeulen N, Parker JN, and Penders B. (2013). Understanding life together: A brief history of collaboration in biology. Endeavour 37, 162–171.

Williams R. (2006). Compressed foresight and narrative bias: Pitfalls in assessing high technology futures. Sci Cult 4, 327–348.

Yin RK. (2011). Applications of Case Study Research. Sage Publications, Thousand Oaks, CA.

Address correspondence to:
Federica Fusi, PhD Candidate
Center for Science
Technology and Environmental Policy Studies
School of Public Affairs
Arizona State University
411 N Central Avenue #750
Phoenix, AZ 85004

E-mail: ffusi@asu.edu

Abbreviations Used

API = application programming interface
IT = information technology
NSF = National Science Foundation
R&D = research and development