



# Sustaining large-scale infrastructure to promote pre-competitive biomedical research: lessons from mouse genomics

A. Mishra<sup>1</sup>, P.N. Schofield<sup>2</sup> and T.M. Bubela<sup>1</sup>

<sup>1</sup> School of Public Health, University of Alberta, Edmonton T6G 1C9, Canada

<sup>2</sup> Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3EG, UK

Bio-repositories and databases for biomedical research enable the efficient community-wide sharing of reagents and data. These archives play an increasingly prominent role in the generation and dissemination of bioresources and data essential for fundamental and translational research. Evidence suggests, however, that current funding and governance models, generally short-term and nationally focused, do not adequately support the role of archives in long-term, transnational endeavours to make and share high-impact resources. Our qualitative case study of the International Knockout Mouse Consortium and the International Mouse Phenotyping Consortium examines new governance mechanisms for archive sustainability. Funders and archive managers highlight in interviews that archives need stable public funding and new revenue-generation models to be sustainable. Sustainability also requires archives, journal publishers, and funders to implement appropriate incentives, associated metrics, and enforcement mechanisms to ensure that researchers use archives to deposit reagents and data to make them publicly accessible for academia and industry alike.

## Introduction

The biomedical sciences are now generating reagents and data in rapidly increasing volumes that require proportional modes of archiving and dissemination [1]. Increased sharing minimizes duplication, makes research reagents more cost-effective, and enables novel and follow-on research as well as independent testing of published results [2]. International agreements (e.g. the 1996 Bermuda Agreement and the 2003 Fort Lauderdale Agreement) and several funding agency policies on sharing data and materials (e.g. the Data Sharing Policy of the National Institutes of Health (NIH) of the US [3] and the Biotechnology and Biological Sciences Research Council (BBSRC, UK) [4]) encourage the non-restrictive dissemination and onward use of publicly-funded research outputs. Some journals also have comprehensive sharing requirements [5,6]. Community-wide resource sharing is facilitated by accessible, stable, and well-funded repositories and databases [1], which are archives responsible for receipt,

maintenance, and distribution of materials and data, respectively. We recognize, however, that there are structural and functional differences, as well as linkages, between archives of biomaterials and of data.

Archives are central to the creation of 'research commons' within which research reagents, data and other outputs are shared in a 'pre-competitive,' often collaborative space [2,7–9]. Research commons are supported by governance mechanisms that promote a cycle of deposit, withdrawal, modification, and re-contribution of materials and data, creating a 'network effect' where the value of an archive increases with use [1]. Elsewhere, we have explored the benefits of and incentives for collaborations in mouse model research [8,10,11]. Here, we explore governance mechanisms that may better promote these activities and sustain archives for the community-level sharing of reagents and data, using research on mouse models of human disease as an exemplar for other research communities.

The use of standardized model organisms in basic and pre-clinical research has been instrumental to our understanding of

Corresponding author: Bubela, T.M. ([tbubela@ualberta.ca](mailto:tbubela@ualberta.ca))

the genome, gene function, and genetic contributions to human diseases. Communities using these reagents have developed archives to enhance sharing and to build a communitarian ethos [1]. For example, key databases have community-integrative effects by compiling, systematizing, and disseminating information on key model species. Examples include The Arabidopsis Information Resource (TAIR) for the plant model thale cress (*Arabidopsis thaliana*) and Mouse Genome Informatics (MGI) for *Mus musculus* [12]. The mouse-model community has well-established norms for sharing infrastructure, use of standardized models, and collaborative ethos and practice [10].

Historically, archives for mouse resources were developed to distribute materials and aggregate data from individually funded projects. For example, in response to community demands, the Jackson Laboratory (JAX) became one of the first animal repositories in the 1930s and was established as a frozen embryo repository in 1979 [13]. The Harwell Frozen Embryo and Sperm Archive (FESA; <http://www.har.mrc.ac.uk/services/biological-services/genetically-altered-line-archives>) was founded in the mid-1970s to protect valuable mouse strains and to distribute mice amongst UK scientists. The current trend, however, is to expand operations of public archives to support large-scale consortia for high-throughput generation of research reagents. These large-scale efforts involve transnational networks of collaborating research institutions, funders, archives, and end users. Archives are key linking nodes in these large networks. For example, the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) aims to link European biobanks in a federated hub-and-spoke model [14,15]. The ELIXIR (European Life-sciences Infrastructure for Biological Information) initiative aims to support the integration and use of life-science research data generated across Europe, with the European Bioinformatics Institute (EBI) acting as a data repository for a series of large international projects, such as ENCODE [16,17]. In mouse functional genomics, archives are crucial to the efforts of (1) the International Knockout Mouse Consortium (IKMC; <http://www.mousephenotype.org/about-ikmc>) to knockout (inactivate by replacement or disruption of DNA sequences) every protein-coding mouse gene and (2) of the International Mouse Phenotyping Consortium (IMPC; <http://www.mousephenotype.org/>) to characterize IKMC mouse strains using standardized protocols [18].

Funding agreements for the IKMC and IMPC directed that resources generated by the projects should be accessible beyond the initial funding term. Accessibility, however, is closely linked to the long-term sustainability of the archives housing and disseminating resources [1,19]. In a commons framework, resource sustainability requires governance supporting broad-based participation in making and using community archives [7]. In addition, legal structures and funding mechanisms must support the transnational and long-term nature of resource-generating initiatives, exemplified by the IKMC and IMPC. Such transnational partnerships have an unprecedentedly broad scope of operations and service and require concomitantly expanded skill-sets, technical development, and budgets. Unfortunately, existing funding and governance structures remain limited in time and jurisdiction, posing a challenge for archive sustainability [15]. Large-scale archiving initiatives urgently require transnational inter-funder policies and formal agreements, long-term funding, and legal

agreements enabling such support. However, research governance remains parochial, 'nationally orientated and based on the "one researcher, one project, one jurisdiction" model' [14: 377]. Our case study of mouse functional genomics identifies governance mechanisms that may sustain archives and strengthen their performance as infrastructures in a collaborative and globally networked research milieu.

Whilst we concentrate in this paper on the archiving and distribution of mouse-related research materials, many of the same problems of sustainability and governance also apply to data resources. Where appropriate we discuss some data resource solutions that are currently being implemented, as potential solutions to the problems of bioresource sharing. There are, however, significant differences, which affect models of governance and funding. For example issues of intellectual property and licensing are now rarely a problem with large data resources, and although there exist subscription models, such as the Human Gene Mutation Database (HGMD) [20], which treats its data as proprietary in a time-limited fashion, much data are currently freely available and are generally curated from public resources. International data distribution and federation are also much easier with databases than bioresources, and many of the major constraints and challenges for bioresource repositories do not exist for data.

Bioresource centres are, additionally, much more subject to the impact of disruptive technologies, which affects long term viability. For example, the BACPAC repository of artificial chromosomes was heavily used until the advance in DNA sequencing technologies reduced use to the extent that the repository has had to change its emphasis to remain viable. The ADDGENE (<http://www.addgene.org>) repository of plasmids, on the other hand, has a sufficiently wide remit, low archiving and access costs to make it one of the most flexible and successful bioresource repositories [21]. In the mouse field the recent development of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) for genome editing [22] potentially provides challenges to resource centre business models and is discussed in depth below.

The following sections describe the background of our case study, our use of qualitative methods, and our key findings on (1) gaps between institutional funding policies and the aims, activities and needs of archives; (2) governance mechanisms and business models to promote archive use and sustain them in the long term.

## Background of the case study – the IKMC and the IMPC

The IKMC was launched in 2007 to generate mutants for all protein-coding genes in mouse embryonic stem cells (mESCs) from the mouse strain C57BL/6N. The IKMC mission was to provide 'a core public archive of ES cell clones on a single, uniform, genetic background, each clone carrying an engineered mutation in a different gene' and 'to extract biological insights from this resource...' [23: 581]. There were several considerations behind the ambitious systematic effort to generate mouse 'knockouts' for biomedical research. First, the mouse is a valuable animal model in biomedical research owing to its small size, low maintenance cost relative to larger animal models, genetic similarity to humans, amenability to genetic modification and analysis, and availability of inbred lines. Second, prior hypothesis-driven knockout studies did not provide sufficient, unbiased, coverage of both genome and

phenome to carry out systematic genome-scale analysis of gene function [24]. Third, before the advent of next generation sequencing, attempts at saturation mutagenesis of the mouse genome with chemical mutagens were slow and expensive, requiring extensive breeding programmes to identify mutations [25].

Key research centres in the United States (US), continental Europe, the United Kingdom (UK), and Canada launched the IKMC effort, using gene targeting and, to lesser extent, gene trapping [26].<sup>1</sup> The IKMC included the KnockOut Mouse Project (KOMP) funded by the National Institutes of Health (NIH, US), European Conditional Mouse Mutagenesis Program (EUCOMM) funded by the EC, and the North American Conditional Mouse Mutagenesis Project (NorCOMM) funded by Genome Canada (GC) [23,24,27,28]. By 2012, the IKMC had generated more than 17,400 ES cell clones, which could be 'readily transferred between laboratories and across international boundaries' [23: 581] and from which large-scale production centres have generated more than 1700 mouse strains, most of them conditional knockouts, frozen as sperm [18,28]. Frozen sperm is a low-cost alternative to embryo-freezing; sperm can easily be transported on dry-ice, avoiding liquid nitrogen or live-mice shipments.

In 2011, the IMPC was established to generate a complete phenotypic profile for mice derived from the knockout ES cells of the IKMC, producing an 'encyclopedia of mammalian gene function' [18]. The IMPC's 'systematic phenotyping' of IKMC mouse strains investigates the pleiotropic functions and effects of genes in physiology and development and how genetic mutations are associated with diseases [29]. Running from 2011 until 2021, the IMPC's high-throughput generation of materials and data, with community-wide relevance, represents greatly enhanced scale, quality, and potential impact compared to earlier small-scale production and phenotyping efforts [30,31]. Aiming for community consensus in its efforts, the IMPC actively elicits inputs from researchers on priorities (e.g. disease area) for knockout mouse production and phenotyping [32: 10]. Through the UK MRC (Medical Research Council) Mouse Network (<https://mrcmousenetwork.har.mrc.ac.uk/>), the IMPC coordinates with researchers interested in specific domains to deliver mice for more detailed phenotyping and assessment.

The IMPC's standardized phenotyping screens cover diverse areas, including behaviour, bone, and muscle development; neurology; vision; haematology; immunology and allergy; cardiovascular and lung function; energy metabolism; and pathology [32,33]. The protocols have been developed from earlier efforts of the Wellcome Trust Sanger Institute Mouse Genetics Programme (WTSI MGP) (<http://www.sanger.ac.uk/science/collaboration/mouse-resource-portal>), and the European Mouse Disease Clinic (EUMODIC) project (<http://www.eumodic.org/>). The EC-funded EUMODIC was the first internationally coordinated, large-scale phenotyping effort and was comprised of four mouse phenotyping centres (MRC Harwell, WTSI, the Institut

Clinique de la Souris (ICS; Mouse Clinical Institute, MCI) Strasbourg, and the Helmholtz Zentrum Munich German Mouse Clinic (GMC)). These centres are founder-partners in the IMPC.

The IMPC high-throughput model requires participating centres to phenotype a minimum of 100 lines annually. Thus partnership is possible between centres with the required capacity to meet targets, that is, with high health status animal facilities, transgenic laboratories, phenotyping platforms, trained staff, and operating funds. Thus the IMPC involves major mouse genetics centres, archives (repositories and databases), live-mouse housing facilities, and leading funders in Europe, the US, Canada, and more recently, China, Japan, Korea and Australia (Table A.1 and Figure A.1). IMPC production and phenotyping centres generate vectors, live mice, and phenotyping data using the IKMC's mESC stocks. Dedicated mouse facilities at production and phenotyping centres provide closely controlled housing for live mice bred from repository strains for phenotyping experiments [34]. Repositories process and distribute mouse materials (sperm, embryos, and tissue) and live mice; databases disseminate phenotyping data, both amongst consortium partners and to end-users through a centralized open-access portal (<https://www.mousephenotype.org/data/search>) [35].

Supplementary Table A.1 and Figure A.1 related to this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.nbt.2015.10.002>.

The participating repositories (e.g. the European Mouse Mutant Archive (EMMA) network and the US-based KOMP Repository) are transnational in scope, serving individual research groups in different countries as well as high-throughput projects such as the IKMC and IMPC. This versatility is made possible by their extensive technical and human resources. Specialized teams employ standard methods to manage archives and colonies, to provide information technology support (collecting, preparing, and disseminating phenotyping data), and to deliver services (handling customer queries, tracking usage, legal agreements, material transfers, finances, communications, and outreach).

The archives' participation in high-throughput resource generation requires coordination across locations and the standardization of production, phenotyping protocols, and data reporting formats. Such coordination and standardization are key to the generation, processing, archiving, and dissemination of data and materials within the consortia and to end-users [35]. Shared standards enable robust integration of activities of transnational centres for mouse strain production and phenotyping as well as data processing and dissemination. This degree of standardization and specialization is in marked contrast to small institutional archives, often run by researchers themselves. Similarly, the live mouse facilities serving the IMPC, with their expanded and dedicated spaces, specialized staff, and enhanced equipment and security, are larger in scale than most small university facilities. Expanded mouse facilities optimize use of time, space, and resources by consolidating animal housing with research spaces under one roof. Centralization of functions and higher-density mouse housing allow researchers to fully phenotype more mouse strains in shorter time periods, achieving higher throughput. Close monitoring of the stand-alone buildings reduces risks of contamination and colony loss [34]. While smaller archives fill a valuable niche

<sup>1</sup> While gene trapping costs less than gene targeting, it is random and not applicable to all genes. Gene targeting works with most genes and, more importantly, can be used to produce 'conditional' knockouts, with effects that can be activated when required in specific tissues or at specific times, allowing greater experimental flexibility and avoiding embryonic or perinatal animal death, common with homozygous knockouts [23,26].

in providing domain-specific resources and support for larger consortia, we focus here on archives that support high-throughput endeavours.

IMPC partners form a confederation through a formal memorandum of understanding (MoU) [32,36]. The IMPC's governance structures are conditioned by the fact that the research and funding partners operate in diverse jurisdictions. While a loosely federated structure may accommodate some jurisdictional diversity, consortium partners may find it difficult to agree on common legal terms for distribution of materials to end-users or even amongst consortium members [11]. Additionally, there is a divide between the transnational aspects of the Consortium and the funding by national agencies for most members. For example, the governments of Australia and Japan support the participation of the Australian Phenomics Network (APN) and the RIKEN Bioresource Center (RIKEN BRC), respectively. Transnational funding is available to relatively few partners, such as the WTSI, which receives funding from the NIH and also the UK Wellcome Trust [32,37]. Indeed the IMPC Business Plan acknowledges that while 'funding for mouse production centres is likely to be granted by individual funding agencies, there is a need to ensure clearly coordinated funding for the centralized informatics activity' [32: 25].

## Methods

Between August 2012 and November 2013, we obtained primary data from in-depth semi-structured interviews of (1) representatives of leading biomedical research funders; (2) managers of repositories and databases; and (3) managers of mouse facilities. In designing the interview guides, we consulted Canadian archive managers and drew on peer-reviewed articles, policy briefs, news reports, and organizational websites. Questions focused on governance and best practices for sustaining archives and strengthening research commons.

### Recruitment strategy and intake

We invited participants on the basis of their publication record, institutional affiliations, and role in the IKMC and IMPC. We recruited 15 archive managers (9 repositories and 6 databases), 6 mouse facility managers, and 8 funders. Participants were located in the US, Canada, Europe, Australia, and Japan.

### Recording and analysis of data

Interviews were 45–60 min long, conducted over telephone. Archive and facility managers described (1) their professional roles; (2) technical aspects of colony and archive management, archiving, and distribution; (3) funding and business models of archives; (4) effects of funding on archive efficiency and services to users; and (5) challenges and solutions in attracting users and funding. Funders described (1) their agencies' roles in supporting research infrastructures; (2) challenges of long-term transnational funding for archives; (3) measures to enhance public deposits of research output; and (4) challenges in enforcing open-access policies. We augmented interview data with documentary analyses.

We anonymized transcripts and organized and analysed data for salient themes, or 'codes,' using NVivo 10© (QSR International) qualitative analysis software and the analytical method of

'constant comparison', involving iterative and linked data collection and analysis [38]. We compared transcripts to identify themes in initial interviews. We then explored these themes in subsequent interviews. We also re-analysed earlier transcripts to incorporate themes, perspectives, and information emerging from later interviews. Coding began with describing basic themes, or 'open' codes (e.g. 'Locations, networks and setup'; 'Funding sources'). Next, we integrated open codes into 'axial' codes describing relationships, interactions, and consequences presented in our results (e.g. 'Large-scale research infrastructure lacks sustained public funding'). We stopped coding when analysis yielded no new themes. We checked quotes to ensure our retention of speakers' intent and reviewed codes for appropriateness. We shared our findings with our informants, four of whom provided critical feedback and refined our analyses, particularly of European initiatives for infrastructure development, implications of novel genome-editing technologies, and challenges of archiving and distributing mouse lines in a federated system.

Below, we present interview excerpts with alphanumeric codes representing quoted funders (e.g. F#1) and managers of repositories (e.g. R#1), mouse facilities (e.g. AF#1), and databases (e.g. D#1). We use square parentheses around inserted explanations [no italics], condensed transcript segments [*italics*], and concealed identities [no italics].

### Ethics

The Research Ethics Office of the University of Alberta approved our study. We protected identities of participants and maintained data in secure university locations. We clarified our aims and methods to participants, who provided signed consent and agreed to the publication of anonymized data.

### Limitations

The small sample size, compared to a survey, and focus on the mouse model community may limit the generalizability of our study. However, our data accord with literature on other model organism communities, and mice are the model most used in biomedical research. Further, the sample size reflects norms for qualitative studies and our analyses reached thematic saturation. While qualitative analyses involve subjective data selection and interpretation, our analyses were validated by expert informants to yield a robust account of issues in archive governance and sustainability.

## Results

We find that archive funding rests on institutional infrastructure funding and user-derived fee-for-service income. Both these sources are themselves dependent on complex contingencies. Institutional funding depends on policies of national funders, which experience other demands for limited funds. Income from user fees depends on level of archive use, which, in turn, depends on the quality of its resources, its governance structures, and legal constraints on operations. Costs of deposit and withdrawal need to be balanced against the scientific value of the resource. To justify long term sustainability, archives need to be used and develop metrics to demonstrate the impact of that use – institutional funders should prioritize only those archives that can demonstrate use and impact.



### *Large-scale research infrastructure lacks sustained public funding*

While large-scale infrastructure needs stable support, funding models are rarely designed to meet that need. Funders may provide some 'seed funding' but no long-term support for infrastructure development. Funders prioritize research project funding over long-term investments in infrastructure and in a constrained funding environment, large investments in infrastructure are seen to divert funds from investigator-initiated research [39].

*While there are various mechanisms to help a resource up, there are not so many for maintaining it. [F#5]*

*Most CIHR [Canadian Institutes of Health Research; principal Canadian funder of academic biomedical research] funding is for project-based research. Large payments to foundations like Canada Foundation for Innovation [CFI; main Canadian infrastructure funding initiative] to build infrastructure put greater burden on agencies covering project-based funding. [F#2]*

As a result, the development and operation of research infrastructure rely on multiple funding sources within an unpredictable public funding environment.

*Resources get core-operating funding from diverse federal and provisional agencies. They have quite the patchwork puzzle to assemble to sustain their core operations. [F#2]*

*The CFI doesn't have a regular funding cycle. It launches new competitions when it receives funds from the federal government. So it's hard for researchers to plan when to submit infrastructure funding proposals. [F#1]*

Archives and animal facilities struggle to secure long-term funding. Short-term funding cycles, geared to shorter project lives, and shifts in national and institutional funding priorities threaten development and retention of expertise, physical locations, and equipment. Shortfalls in these areas, in turn, affect quality of operations and services.

*Most funding cycles now are the relatively short term project-based model. In that scenario, even five years is very long. A resource needs consistent core funding for a decade plus to ensure its capacity to send data or materials to researchers. Repositories need that classic aid-based funding from some government-related organization that commits to developing and maintaining a resource without time limit on the funding, which would continue so long as the resource justifies its need. [F#2]*

*The number of mouse strains that we maintain has been increasing annually. We will need to renew [facilities and equipment]. However, government support is decreasing across Japan. Our challenge is to minimize the rate of this decrease by negotiation with the government. [R#5]*

*Our repository infrastructure is 100 per cent subsidized by [research institution]. As budgets change, there is no guarantee of continued support. [R#2]*

Large archives inevitably have higher maintenance and operational costs than small-scale equivalents, although there are economies of scale for the system as a whole. Salaries dominate budgets because skilled personnel are essential to maintain service quality. Funding shortfalls limit the availability of equipment and personnel, such as liquid nitrogen for cryopreserving mouse strains and personnel for monitoring colony health. Funding shortages or decreased usage necessitate hard choices between workforce reductions that affect service quality and compensatory price increases that may discourage users.

*Being a larger facility with specialized staff and equipment, we can take on projects for researchers in other locations in Australia. But there's also a high cost with [specialized] infrastructure. [AF#2]*

*Funding affects our staff continuity. It takes six months to a year to get new staff trained and comfortable in their roles. The IMPC grants are short term. If these funding streams don't come through, we have to lay staff off. That affects our capacity and throughput. [R#3]*

*Replacing skilled staff can take several months. Backup persons could continue production but it would take three or four months instead of two. [R#1]*

*If we don't get enough funding or if we can't sell enough materials, we will need to either increase the prices or reduce the workforce. We will probably still have the same materials but with a different price or reduced workforce, it may take longer to fulfil services and sales. [R#1]*

### *Archives operate transnationally; funding is national*

The nationally bounded nature of funding policies and priorities is a problem for large archives with transnational services. Funders find it difficult to harmonize policies and priorities to support archives in distinct jurisdictions but with networked operations.

*The emerging thing is work in an international consortium model. But we're challenged to contribute funds to repository efforts outside the country when scientifically and administratively it might make sense. We have had discussions around cross-funding agreements to support these platforms. There is no formalized process in place, but that conversation is increasing. [F#4]*

Developments in Europe hold some promise of sustainable funding models for distributed research infrastructures. In 2002 the European Strategy Forum on Research Infrastructures (ESFRI; [http://ec.europa.eu/research/infrastructures/index\\_en.cfm?pg=esfri](http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri)) was launched, bringing together representatives of EU Member States and Associated States, and the EC, to support coherent and strategic policy-making on European research infrastructures. Another ESFRI objective was to facilitate multilateral initiatives towards the better use and development of European

and international research infrastructures. New governance models from the ESFRI aim to address the limitations of short-term project based funding of research infrastructures. ESFRI prioritizes infrastructure projects for support and publishes its recommendations in ESFRI Roadmap reports. Prioritized projects are eligible to apply for EC funding in response to specific calls.

The first ESFRI Roadmap included 35 infrastructures across all scientific disciplines, among them, the INFRAFRONTIER infrastructure, which aims to increase capacities for production, archiving, distribution and phenotyping of mouse models [40]. In a preparatory phase a business plan was developed for the INFRAFRONTIER research infrastructure and led to the formation of a transnational legal entity funded by stakeholders. On 11 April 2013, INFRAFRONTIER, centred at the Helmholtz Zentrum Munich, was incorporated to acquire the status of a German private limited company (GmbH; <https://www.infrafrontier.eu/infrafrontier-research-infrastructure/organisation/infrafrontier-gmbh>). The formation of the INFRAFRONTIER GmbH is an interim measure towards further development into a European Research Infrastructure Consortium (ERIC), a dedicated legal entity developed for research infrastructures in the European Union [41]. Notable ERICs include the BBMRI ERIC (<http://bbmri-eric.eu/>) in biobanking, and the European Advanced Translational Research Infrastructure in Medicine (EATRIS; <http://www.eatris.eu/index.html>) ERIC, in translational medicine. The stakeholders of ERICs are states (not institutions) committed to stable funding of infrastructures.

As INFRAFRONTIER coordinator, the Helmholtz Zentrum Munich was awarded project funds from the national research ministry to develop INFRAFRONTIER nationally, cover funding for the GmbH shares, and to contribute to the IMPC. While there is yet no EC funding for large-scale phenotyping efforts, and all IMPC contributions remain nationally funded, the EC-ESFRI process and establishment of the INFRAFRONTIER legal entity has helped a few EU archives and mouse clinics to obtain visibility and funding. Nevertheless, there remain challenges, notably with persuading government agencies to provide long-term funding for infrastructure initiatives. It is also debatable whether the ERIC model can be successfully exported to non-EU settings.

*EMMA is a network of different national partners, who all have to go back to their ministry for funding. Ministries are used to giving project funding for four years. This INFRAFRONTIER process is transitioning from project funding to co-funding, to get stable funding. INFRAFRONTIER is used on a national level to trigger additional co-funding and investment. That is a challenge because the administrators are used to project-based funding and don't want to commit themselves to giving you a couple of millions per year over 10 years. [R#6]*

There are relatively few successful examples of hybrid and transnational models of funding for research infrastructures. The Universal Protein Resource (UniProt) (<http://www.uniprot.org/help/about>) for protein sequence and annotation data receives most of its funding from the NIH, with additional contributions by the Swiss Federal Government for the UniProt partner Swiss Institute of Bioinformatics and the European Molecular Biology Laboratory (EMBL; <http://www.embl.de/>) [42]. The Worldwide Protein

Data Bank (wwPDB; <http://www.wwpdb.org>) is a collaboration of major protein data banks and repositories in Europe, Japan, and the US, supported by numerous funding agencies [39].

### *The need to develop viable revenue-generation models for archives*

Our participants debated optimal revenue-generation models for community archives. While funders insisted that the research community needs to develop long-term plans for archive sustainability and revenue-generation, they offered no specifics for developing such plans. Currently, most repositories charge for products and services, with many attempting a cost-recovery model, which generally requires long-term subsidies from leading funders such as the EC for EMMA and the NIH for the KOMP Repository. Deposits of mouse-related research materials into EMMA are currently free of charge, with costs subsidized from EC funding. In contrast, in Canada, a repository manager described how lack of funding limits distribution. Moreover, the repository's ability to derive income from deposited strains is limited by restrictions on onward distribution imposed by the original depositors.

*EMMA deposits are free and users have to pay to get the mice out. But the deposits are only free because the EC funds all deposits. EMMA have funding to pay for their staff, and the same staff deposit and distribute. In Canada, we have no funding except potentially through research grants. That just pays for deposit of lines by the grant-holders. Unless we have ongoing funding the resources aren't accessible because we need staff to pull it out of the freezer and send it off to people. We know if we are not getting income from some strain. . . most of the lines deposited are not for distribution. Occasional withdrawals don't sustain a repository. [R#2]*

In the IMPC, the ES cells are concentrated in specific repositories. The federated, international nature of the resources raises additional challenges for distribution and income.

*Only two or three repositories have the ES cells. UC [University of California] Davis [lead partner] of the KOMP Repository has all the ES cells produced by the KOMP program. The Helmholtz Center in Munich has all or most ES cells produced by EUComm. Sanger has a large number produced by EUComm and also by the KOMP program. And NorComm has produced a few. [R#7]*

As costs of archiving, production, and quality control differ between sites, charges paid by users also differ, for example, between the KOMP Repository (<https://www.komp.org/fees.php>) and European Mouse Mutant Cell Repository (EuMMCR) (<https://www.eummc.org/faq#handling-fee>). Repositories distributing materials to for-profit users (e.g. the KOMP Repository) charge those users additional licensing fees. There are obstacles to the distribution of less-requested and unique lines from smaller repositories that may receive fewer user requests and lack the staff and infrastructure of their larger and better-known counterparts. As more nationally funded partners come on board, business and funding models remain parochial, with no clear solutions for distributing collections across partner archives.

*Collections in smaller repositories that aren't popular or not duplicated at multiple repositories, or both, may be lost as there isn't any mechanism for transferring collections between repositories and very few national funders may want to allocate budget to stock foreign repositories. [R#2]*

Within the EU, the EMMA network seeks to minimize internal competition by adopting a standard fee structure for distribution, notwithstanding the variable costs of archiving and production amongst EMMA partners (<https://www.infrafrontier.eu/procedures/emma-repository/emma-service-fees>). A standard fee structure is not easily achieved 'given the quite different fixed operating costs and levels of subsidization in different geographical locations' [R#2]. Moreover, in the EMMA case, users still bear shipping costs. Variable, often considerable, shipping costs can lead to users 'shopping locally' [R#6], reinforcing the trend of parochialism that runs counter to the ethos of the mouse commons.

A challenge to distribution was the delay in EUCOMM's ability to distribute ES cells to for-profit entities [11]. The delay was partly due to potential intellectual property (IP) liabilities arising from the mode of generation of the high-throughput resource, utilizing multiple reagents and methods that may be subject to third party IP rights. In the US, in contrast, an Authorization and Consent provision in the funding agreement with the NIH, enabled by the *Federal Acquisition Regulations* (FAR), in practice, immunized KOMP members, as federal government contractors, from potential patent infringement litigation [43]. Thus, the KOMP repository is free to distribute materials to commercial users. This provision is peculiar to US Federal Government contracts, discretionary, and unlikely to be implemented in other jurisdictions. However, on 2 July 2014, EUCOMM partners were able to address their distribution challenge by licensing commercial mouse-model developer genOway [44], which held or had secured licenses to third party IP, to distribute EUCOMM resources to commercial users. The EUCOMM-genOway agreement allows genOway to provide commercial users with the rights to use EUCOMM's existing archive of conditional knockout models. In addition, users have a defined time frame in which to access EUCOMM materials and generate their own knockouts.

Potential patent infringement is a multi-layered issue [10]. First, public repositories and their academic users are open to legal action. For example, in 2011 the NIH intervened with 'Authorization and Consent' to protect JAX from a February 2010 lawsuit by the Alzheimer's Institute of America (AIA), a non-practising entity [45]. The AIA had filed suit against JAX for allegedly profiting from the distribution to academic researchers of mouse models carrying the 'Swedish mutation' associated with early-onset Alzheimer's disease. The AIA held a US patent for the mutation [46], which has since been invalidated [47]. The NIH intervention not only shielded JAX from rent-seeking litigation but also ended the AIA's settlement-related demands that JAX divulge identities of researchers who had received the mouse model and could have been exposed to AIA lawsuits [45,46]. Second, commercial entities purchasing research reagents carrying infringement liabilities also become targets for litigation. Limits on distribution to for-profit entities has implications for public repositories, which generally charge for-profit users fees that are higher than for academic and non-profit users, a tiered pricing model that yields somewhat higher returns.

*Commercial entities can access our lines under the same MTA as academic or non-profit researchers. The difference is that the academic customers pay less. We have obligations to the holders of the patents on the technology that we use to generate our library. When we make a commercial sale we have to pay them royalties from that. [R#1]*

#### *Funders' responsibility for long-term support to archives*

Repository managers emphasized that funders need to recognize that their support to repositories represents large-scale cost savings and efficiency gains. Archives enable researchers to spend their constrained research funds on investigations rather than on making reagents, housing live animals, and distributing them to colleagues.

*[One argument to make to funding agencies is] that we're really saving valuable resources that your funds have helped pay for and we're saving researchers grant money because they don't have to keep the animals as live mice. [R#2]*

*Before these resources were developed, the individual investigators would have had to spend large chunks of their funding on creating their own resources. They can now spend most of their funds on scientific experimentation, testing hypotheses and achieving outcomes. The resources mitigate current financial constraints for hypothesis driven science. [R#7]*

Some funders were sceptical of providing long-term support to archives, in particular, to those whose use declines as needs and priorities change over time.

*We can't always be the ones sustaining resources. Some resources are all the rage and then go out of fashion. We need honest conversations about whether a resource is something we're hanging on to just because we invested a lot in it, and we're unsure if we've extracted all the value from it, or whether it is essential and should be maintained. How much money or effort to keep it alive long enough that it could re-emerge as something useful versus just letting it die completely? [F#4]*

However, repository managers felt that even successful repositories will still require supplementary funding support. For example, while the successful KOMP repository no longer depends on NIH funding, sustainability remains a concern because of the costs associated with maintaining user satisfaction.

*The KOMP repository, NIH funded for four years, is no longer funded by the NIH. It needs to rely solely on income derived from distribution of products to the research community. This has been a highly successful project, the first and only NIH supported resource to go completely self-sufficient. However, the repository needs to continue to provide customer support, technical support, and a website that is informative, helpful and useful so that people continue to want to obtain these products. These are absolutely necessary to derive income in order to maintain*



*the archive. So being self-sufficient does not mean that the archive is easily sustainable in the long run, for decades to come. [R#7]*

Repository managers emphasized that few repositories can become self-sustaining purely on the strength of distributional cost recovery, even taking account of economies of scale [19]. Income from high-demand strains may not be sufficient to subsidize maintenance of strains in lower demand but with untapped potential. Maintaining little-used strains is an important role for a repository because research trends are unpredictable, and such a strain may become an important resource in the future. External funding therefore remains important to storing such strains; this task falls on public repositories because it is not viable for commercial repositories to maintain unpopular strains. For example, there was a resurgence of interest in lines generated from the large-scale, worldwide Ethyl Nitrosourea (ENU) mutagenesis programmes of the early 2000s. While thousands of ENU lines, each unique, were archived, the difficulty of mapping candidate mutations led to many lines being 'left on ice'. Recently, however, next generation sequencing techniques (NGS) have facilitated mutation identification. Many centres are now re-sequencing archived lines with a surge in new disease models, and importantly the discovery of allelic<sup>2</sup> series of mutations in genes already implicated in diseases [48].

#### *Obsolescence: new technologies and reagents may alter use of repositories*

Shifts in the community's use of reagents (e.g. from gene trap lines to conditional knockout strains) and model organisms (e.g. from mice to rats) either have limited, or may limit, use of mouse resources.

*Our repository only has gene trap lines, which are out of favour right now. Most customers prefer conditional knockouts because you can model closer to what usually happens in vivo. You can direct effects in specific tissues and tailor your experiments to your area of expertise. [R#1]*

*Mice are cheaper than rats, but rats are bigger animals. So you can take larger samples from rats. The rat's physiology is more akin to the human. But we don't have the transgenic technologies for rats that we have for mice. If rats became more amenable to genetic manipulation then our archive's activity would diminish quite substantially. [R#3]*

The recent emergence of genome editing technologies may present a new challenge to the viability and sustainability of archives, especially those maintaining and distributing ES cell lines. The technologies include Zinc Finger Proteins (ZFNs), Transcription activator-like effector nucleases (TALENs) and most recently, the CRISPR/Cas9 system (Clustered Regularly Interspaced Short Palindromic Repeats and their CRISPR Associated protein 9) [49]. Accessible in any laboratory with molecular biology experience and some animal expertise, these technologies enable ma-

nipulation of genes in simple and complex experimental organisms (including bacteria, mice and primates) and may allow genomic alterations directly in embryos. Thus, researchers can avoid the lengthy and unpredictable process of genetically modifying ES cells, developing them into embryos and then into adult organisms, which may or may not be able to transmit the alterations to offspring. Some participants suggested that this advantage of genome editing technologies could reduce reliance on ES cells and their archives.

*With ES cell manipulations, you still have a barrier to your success and that is whether that ES cell can actually contribute to the germline ['germline competence', or ability to successfully contribute to gamete formation and transmit targeted genes to progeny [50]]. Vertical transmission is quite a hurdle. Now these other technologies are able to bypass the ES cell and directly modify the genome in the embryo ensuring transmission. [R#7]*

*Precision targeted genome editing of mice in vivo will mean that it's no longer necessary to manipulate a stem cell to get the mouse you need. We have lived off mouse stem cell banking for a long time. That's changing. [D#3]*

*While the KOMP repository distributes both mice and ES cells, the EUComm repository, EUMMCR, only distributes ES cells. This has implications for sustainability ... with the advent of CRISPR/CAS9, the demand for ES cells could decrease. [R#2]*

An alternate view was that the new technologies, with limited targeting efficiency and relatively high associated costs, posed no significant or immediate challenge to ES cell archives.

*We have been able to work with ES cells in a very high throughput fashion. With those new technologies, that hasn't been done satisfactorily yet. I don't believe that CRISPRs, TALENs, Zinc Fingers will reduce the value of the current KOMP, EUComm and NorComm resources. [R#7]*

*To make a mutant mouse from scratch with CRISPR/CAS9 and this genome editing technology may never get as cheap as it is to pull a stem cell out of a freezer and then breed it up. [D#3]*

Despite enthusiasm for genome-editing technologies, off-target effects (unwanted mutations elsewhere in the genome) are a serious issue and may make the reagents irreproducible [51,52]. Participants were concerned that mice made 'in house' and not distributed through repositories would not be subject to verification of characteristics such as background strain, disease status, and mutation identity, all tested by repositories. Interactions amongst genetic background, disease status, and phenotypes are amongst the most common reasons for failure of the same mutation to generate the same phenotype. Where phenotypes are, for example, weakly penetrant or subject to modifiers, mice with the same edited mutation may give rise to disparate phenotypes for often completely unknown reasons [53].

<sup>2</sup> An allele is a form of a gene, found at a specific chromosomal locus.



In the aggregate, the increased use of individualized ‘cottage industry’ methods has several potential negative effects. It could divert already scarce research funds to the costly and piecemeal task of making reagents, which may have diminished quality and standardization in comparison with reagents produced using standard protocols in large resource-making efforts. Second, there are potential costs and losses from storing those reagents in ill-monitored and unstandardized small institutional freezers associated with contamination and damage.

*The mouse and human genomes were sequenced not by single laboratories doing part of a chromosome each but by laboratories working in consortia in high-throughput projects. If someone said, “I’ll do this gene using ZFNs, and I’ll get my mouse and I’ll study it,” they’d be returning to the Dark Ages, when people were knocking out one gene at a time in their individual labs. . . We need improved efficiencies on those new technologies to enable them to be high throughput and then allow some big laboratories to use those technologies to create new mutations, and then make sure those resources are available to the broader community. [R#7]*

*It’s easier to standardize in big science because it is an economy of scale. People have to standardize a lot of their stuff because they have to do the same thing repeatedly. If you’re running a small institution with a few mice and small experiments, you would not have needed to make an investment to large-scale standardisation or infrastructure. [D#1]*

*The university labs haven’t got the infrastructure or staff to do the health monitoring which we have. [R#3]*

Research groups deposit individualized reagents in archives at a lower rate than the high-throughput projects because groups are not mandated to do so, or because the quality of the reagent is not high enough to pass the quality control standards of the repository. Lack of deposit returns the community to the ‘bad old days’ where materials and data were not shared, behaviour that ironically led to institutional policies on data and materials sharing and funder commitments to generation of high-throughput community resources and support for archives. Lessons learned are that widely dispersed resources lead to increased direct and marginal costs for funding agencies for the distribution of reagents associated with publications, duplication of resources, and increased mouse usage. The latter two effects counter the ethical experimentation aims of reduction, replacement and refinement, or ‘3R’ [54]. Thus, these novel technologies may be disruptive not only to archives but also to norms of open, reproducible and ethical science.

#### *Promoting the research community’s use of public archives*

To be sustainable, archives must be valued and used by their respective communities. Archives provide insurance against loss of reagents and data, enable ethical maintenance and distribution of experimental organisms, and facilitate distribution to third parties. Our participants, however, described the under-use of

archives as a serious challenge to their sustainability and offered some solutions to enhance usage. Use in this context means both deposit to, and withdrawal from, archives.

#### *Challenges to the deposit of reagents and data*

At the community level, concerns over publication priority [55] and commercial interests in materials and data may inhibit deposit or else prompt depositors to restrict the onward distribution of materials by archives to third parties, especially for-profit entities. Privacy and patient consent for human reagents are additional concerns for deposit and use of biobanks.

*When you start dealing with commercial interests or with academics working with commercial interests, we can tell them that we need data returned, but it won’t happen. Also, I’ve dealt with projects where the researchers were hesitant to deposit sequence data because they weren’t done analysing it. But with such data you’re never done analysing. At some point you need to let others look at it too. [F#3]*

However, practical impediments to deposit of data and materials also exist, at the end of the depositing researchers and of the receivers, that is, journals and archives. Archives generally require depositors to perform pre-submission quality control of reagents and data. Depositors may view such checks as onerous and avoidable investments of time, personnel, and funds. Database managers also emphasized that researchers need to provide sufficient or well-annotated metadata along with their data. Metadata is valuable background information about the provenance of study aims, materials, and methods. By capturing the ‘tacit expertise’ in data production, metadata enables other researchers to critically scrutinize and interpret experimental results and perform comparative research using those reagents and data [56: 223].

Standard guidelines and quality control procedures exist for the preparation and deposit of metadata. Thus, the project Minimum Information about Biological and Biomedical Investigation (MIBBI; <http://mibbi.sourceforge.net/legacy.shtml>) recommends that researchers standardize descriptions of their experimental protocols to ensure that other researchers can interpret these protocols after accessing them online. The research design of a small academic project may not require the organized compilation of metadata, and researchers may not see a need to process metadata for archiving. However, background information on reagents acquires greater importance in the context of collaborative research, wherein large datasets and the means and processes of obtaining them need to be scrutinized, compared, and verified amongst research groups in diverse sites. In the post-genomics era, metadata acquires increased importance as data circulates across increasingly larger and more diverse networks [56].

*If you just buy a mouse and you’re interested in doing an experiment and data integration isn’t your primary interest, the strain background of the mouse is not part of your experimental design. It is just something that comes with the mouse knockout that you buy or that you make. But if your experiment compared a mutant mouse strain with a wild type mouse [standardised, but not genetically manipulated], it would be important to know both strain*

*backgrounds. If you wanted to do a large scale phenotyping project, comparing mice across centres would be something that highly motivated you. You would want to integrate a mouse phenotype with information on the allele, information on strain background and information on any mapping to a human disease if the mouse was a disease model. [D#1]*

However, the heterogeneity of the research community, with diverse and changing experimental methods, instruments, and reporting formats, precludes consensus about what metadata should capture [56]. Metadata may be layered and complex, which challenges current options for display in journals.

*The journal model doesn't support the complexity of the data that we need to re-do or to fully understand the experiment. Tables are just static representations of some of the data. If data only comes to you from publication, it is difficult to analyse the effects of extra variables that come from the experimental design. [D#1]*

The non-availability or lack of capacity of relevant archives is also a significant challenge to deposits. While funders call for public deposit of research output, funding shortfalls may lower the actual availability of appropriate archives where output can be sent. Also, the large volumes of data and materials being generated are outstripping the available funding and storage capacity of existing archives. In some cases, there are no funds to add capacity to old archives or create and administer new archives for the deposit of novel data and materials.

*We went from the second to the fourth generation of sequencers in about four years. That is accelerating the research but the computing power and storage are getting limiting. There is a value in keeping data because you can find secondary and tertiary uses for it, but how do we keep it? How long? Who pays for it? We can't keep growing the storage. All the National Science Foundation [NSF; US funder] grants have conditions like open access, data management plans, how the data will be preserved and archived. Everybody is struggling on how to meet the NSF regulations right now. [F#1]*

#### *Withdrawal of reagents and release of data – operational delays and the challenge of managing user expectations*

Archives need to be used not only for deposits but also for onward distribution of data and materials to third parties. Our interviews indicated areas for improvement in (i) the technical standardization and quality control of mouse strains and associated data; and (ii) standardization of legal terms and conditions for distribution through use of different forms of licenses ranging from simple conditions of use to material transfer agreements (MTAs). The lesson from the latter is to keep legal terms as simple and as standard as possible to prevent delays due to institutional negotiations over terms of withdrawal and use. Issues regarding the role of MTAs in archive governance are elaborated elsewhere [11,57]. Here, we focus on factors in technical standardization that influence withdrawal of reagents from repositories.

In the case of the IMPC, resources are being accessed either on the basis of a known gene or on the basis of an interesting phenotype, related to the research being carried out by the investigator. The use of reagents such as mice, ES cells, and vectors is connected to the quality of data on, for example, genotype, background strain, and phenotype of the reagents. Access to these data provides users with information on the contents of archives so they can make informed decisions about withdrawals. Databases thus have a significant mediating role in the uptake of materials from repositories, and often 'database curators have to manually align information about each strain of mutants available in stock centres with the online data actually available in relation to those strains' [58,34]. For an end-user attempting to locate a mouse strain and associated data, such consolidation offers a remedy to the current situation for mouse repositories, wherein collections are both scattered and overlapping. In some cases collections are mirrored for biosecurity, in other cases a single strain may only be found at one site. Many valuable mouse strains remain with individual laboratories and institutions. In that situation, databases and search portals (e.g. International Mouse Strain Resource, IMSR; <http://www.findmice.org/>) facilitate access by offering standardized information on the location and technical aspects of mouse strains, stocks, and mutant ES cell lines. However, even with extensive standardization of search terms and nomenclature associated with materials, users may have variable understandings and misplaced expectations of IMPC output and the related repository holdings.

*With these complex IMPC alleles, folks don't always understand the nomenclature and what they will get. They may want the conditional line when in fact we are distributing the knockout line that can be converted into a conditional, which involves an additional breeding step. The first hurdle is to make sure that the client understands what genetic background the mouse is on. Many don't understand how versatile these alleles are or how one actually gets from the initial knockout to the conditional status. [R#3]*

Archive managers also described delays that they felt were inherent to the processes of preparing diverse forms of data for release or for preparing reagents for shipping to customers. They described the unfavourable reactions of some users to the inevitable and prolonged waiting period entailed by quality control and standardization.

#### *Solutions to enhance deposits to archives*

Participants described various incentives to enhance deposit. Deposits that accord with the requirements of a funding agency or journal could be an allowable grant expense [8]. Publishers could expand available options for publishing, using, and citing pre-publication data and experimental metadata [59]. Better markers for proof of deposit, such as standardized accession numbers, could be used as productivity measures for professional advancement, 'if, within your institution's advancement criteria data, depositing data has a beneficial effect on your annual review' [F#3]. In bioinformatics, an accession number is a unique identifier assigned to a piece of data, for example, a sequence on its submission to an archive, for example, GenBank or the European

Nucleotide Archive. While submitted sequence data can be updated, the accession number remains constant [60]. Accession numbers can be provided in publications, providing proof of deposit and enabling other researchers to locate, use, and cite the resources in a manner analogous to previous research publications. Accession numbers are already used, for example, to prove deposit of sequence data in GenBank prior to publication [61]. Citations are a key performance metric for academic researchers and the prospect of getting such recognition of deposits may incentivize depositors.

Archives also offer practical incentives to promote deposits. Many researchers prefer to complete the peer review of their research articles before submitting the relevant reagents and data to public archives for community use [55]. Thus, some resources offer moratoria on reagent and data release: *'a grace period just in case somebody is still working on a publication, at least two years where the line is already archived but not yet visible to the external world'* [R#4]. Some archives incentivize deposits by offering conditional refunds (e.g. the 'Sharing Plan' of the KOMP Repository; <https://www.komp.org/sharingplan.php>) and by absorbing shipping costs and offering credits on future purchases (e.g. the RIKEN BRC; <http://mus.brc.riken.jp/en/deposit>).

Some participants suggested that incentives to deposit, while essential, need to be complemented by the enforcement of data and materials sharing policies. Funding agencies could withhold funds for lack of adherence. For example, from November 2012, the NIH has begun to seriously enforce its policies regarding open-access publication and data deposit. Similarly, in March 2014, four UK funders announced that, from 2016, the Research Excellence Framework, a key audit for research funding, would consider only open-access papers in online institutional archives [62]. Evidence suggests that enforcement actions have raised the percentage of papers placed in the NIH-supported PubMed Central database for public access no later than a year after publication from 75% in 2012 to 82% in 2014. Similarly, The Wellcome Trust's compliance rate rose from 55% in 2012 to 69% in 2014 [62].

*Lately NIH has been really cracking down and you won't get your grant renewal if your manuscripts aren't open access. When NIH do renewals they check if the PIs have published and PIs have to provide the PubMed Central IDs [PMCID] for their manuscripts<sup>3</sup> [F#7]*

Some funders expressed reservations about the feasibility of enforcing sharing policies. While funders carefully review sharing plans at the grant-application stage, they lack resources to monitor actual deposits at project completion. The rapid pace of research inhibits monitoring and enforcement. In many cases, quite reasonably, data may not be published until a year or more after the end of the grant; continuous monitoring for years after the end of a project is unrealistic for funders. Additionally, monitoring of output in 'investigator initiated' academic research is less stringent than that observed in larger projects, such as the high-throughput IKMC and IMPC, which operate under more strictly defined time-

lines, deliverables, and governance structures. Archive managers, on the other hand, insisted that funders, in association with publishers, have the gatekeeping power to track and enforce compliance with data and materials sharing policies.

*The funding agencies and journals have to start saying you must, as a condition of receiving funding and or publishing work, deposit research tools to a repository for anyone who wants to access them. [R#2]*

Archive managers were also keen for journals to enforce consistent attribution of archives in publications. Attribution may involve a direct identification of the source archive; or it may be indirect and guide users to the archive through published accession numbers. Metrics derived from such attributions enable objective assessment of archive usage and role in stimulating research. Archives could use these metrics as evidence for their funding requests. Unfortunately, archive managers lack funds, personnel, and time to track or correct attributions in publications. Archive managers were resigned to attributions being omitted, suggesting that omissions were due to word limits in journals or to the superfluity of citing well-known reagents.

Some funders suggested that community education and encouragement to deposit, instead of enforcement, could generate cultural changes conducive to sharing via public archives. Stringent enforcement of sharing policies could drive a wedge between funders and the research community. As members of funding review panels are often themselves active researchers, they may be reluctant to discipline their peers.

*Each of CIHR's 13 institutes has a scientific director with an academic appointment. So they are still 50 per cent a researcher. This model keeps funders embedded in the research community, working closely with it, not divorced from it. You want to maintain a good bidirectional flow of information so that you remain well-grounded in the needs and capabilities of the research community, while gently instituting policies that are in everyone's best interest. You want community buy-in, not to be imposing rules. [F#2]*

Archive managers also saw outreach to the community as key to improving sharing practice and archive quality and use (e.g. TAIR's elicitation of user input to improve its search and visualization tools) [63]. Unfortunately, only larger and better-funded archives can employ personnel for large-scale outreach.

## Discussion

Effective governance of archives requires considerations of sustainability and remit, that is, that operations meet the objectives of resource creators, funders, and users. Here, we discuss implications of our findings and suggest lessons for other communities (e.g. communities using non-murine models, or initiatives in biobanking), struggling to develop, manage, and sustain research archives. We discuss challenges and solutions for the governance of two levels of archive activity (1) housing and distribution of research reagents and data from large-scale, transnational, resource initiatives; and (2) deposit and withdrawal of resources generated by individual research groups. Many archives support both levels of activity.

<sup>3</sup> Peer-reviewed articles in the NIH-supported archive PubMed Central (PMC; <http://www.ncbi.nlm.nih.gov/pmc/>) are assigned a digital identifier, the PMCID, which NIH-funded researchers can cite to demonstrate compliance with the NIH's open-access publication requirements.



### *Governance of archives to support high-throughput resource generation initiatives*

The publicly funded IKMC and IMPC resources have been generated with considerable effort and expense at a time of increased financial strain for community archives [1]. The resources provide high-quality and accessible tools for basic and clinically relevant research where animal models serve as surrogates for human disease and are key tools for proof-of-concept studies. Wide utilization of these and similar resources would justify past, ongoing, and future investments in making and disseminating them. Resource uptake on a global scale will be influenced by the ability of the archives to operate at the required efficiency, with implementation of appropriate legal and technical standards for resource development, maintenance, and distribution. Our interviews indicate, however, that mouse model archives require, but also lack, stable funding to facilitate their long-term transnational operations and their role in supporting the global mouse research commons.

Funders are limited in their ability to support research infrastructure beyond the standard five-year funding cycle and national jurisdictions. An emergent priority is therefore harmonized and concerted action by national funders to support research infrastructures with transnationally distributed resources and operations. EMBL, ELIXIR [16] and, more recently, Infrafrontier [40], illustrate the implementation of some novel co-funding solutions for research infrastructures. In all three cases governance mechanisms are critical to the success of the infrastructures. ELIXIR and Infrafrontier have adopted quite different structural and legal models with interesting implications not only for these infrastructures but also for the viability of new or un-associated resources in the communities outside their umbrellas.

EMBL, 40 years old in 2014, has an annual budget of around €200M of which ~50% is obtained from member and associated member states on the basis of proportion of Net National Income (NNI), in addition to ad hoc and special donations. EMBL is funded under an international inter-governmental treaty complemented by bilateral treaties, with tax and legal status implications, between EMBL and the countries that host its main laboratory and its outstations such as EMBL-EBI in Hinxton UK. Following the ESFRI initiative, Member States agreed to create ELIXIR, a dedicated initiative to support the coordination, integration and sustainability of Europe's life science data resources. The foundation for ELIXIR is the ELIXIR Consortium Agreement [64] and the Consortium has adopted the legal personality provided by the EMBL treaty. It is now formally a special project of EMBL [65]. ELIXIR Hub funds are accounted for by EMBL, but EMBL does not control spending or governance of ELIXIR and as such the entities remain independent of each other. National contributions to the ELIXIR Hub in Hinxton are based on NNI and are used in a variety of ways to support the network, as described in ELIXIR's recent financial plan [66]. The Nodes, including EMBL-EBI which acts as the "European Node", are principally funded through national sources (in the case of EMBL-EBI, EMBL direct funding) and additionally raise external grant funding for their sustainability. The ELIXIR Hub provides, amongst other services, training, strategy development, resource integration, interoperability standards, and support for the identification of sources of funding and coordination of funding applications. Resources within ELIXIR

may benefit from funding from the Hub for development and integration through "pilots" (<https://www.elixir-europe.org/about/pilot-projects>) and in future "commissioned services", where Nodes receive longer-term support to develop or maintain services. Additionally, a set of databases will be highlighted as "core resources". This will apply to a small subset of what are considered to be the most important databases for life science users. They would be the focus of ELIXIR's policy actions should major sustainability issues arise.

ELIXIR's legal identity allows it to compete for funding, for example from Horizon 2020 from which it has been awarded a major implementation grant starting in September 2015, illustrating how the coordination of national communities and resources can help leverage additional funding at the EU level. ELIXIR thus supports sustainability and development of the network of resources in Europe without the intention of acting as a source of long term direct funding for the operation of individual resources, at least as currently envisioned.

The ELIXIR framework does not, however, offer the possibility of support for resources outside its national Nodes, and adoption as ELIXIR "named services" or eligibility for funding through pilot actions. Any new resource will need to comply with ELIXIR quality standards to be accepted as an ELIXIR named service, and it will need to be associated with national ELIXIR Node and nationally funded. This may be difficult for many useful new resources, either because the originating groups may not be included in a national ELIXIR Node, or lack of sufficient local funds to support the ELIXIR quality criteria. This could have adverse implications for the funding of novel and possibly key data resources from institutions outside the ELIXIR umbrella, and makes national Node governance and national science funding policy critical in supporting the richness of the data ecosystem necessary for a vibrant scientific environment. The risk that national Nodes remain exclusive, or that national funding agencies might use ELIXIR membership as a criterion for local funding decisions, for example, could seriously compromise the development of cutting edge science. This issue is discussed at more length in Attwood et al. [67]. Applying an ELIXIR model to bioresources may result in similar problems, squeezing out the funding of small and niche resources that might be crucial to particular, or novel, fields. The alternative however would be transnational funding of individual national centres, but the potential ceding of control of a national resource to an international organization may be unacceptable to some countries and rules of governance would be critical to viability.

The Infrafrontier model is based on a memorandum of understanding signed by Germany, France, Czech Republic, Finland, Greece and the European Molecular Biology Laboratory (EMBL) for the coordination of the pan-European activities of the research infrastructure. Currently Infrafrontier has legal recognition in Germany as a GmbH, but will in future use the ERIC legal instrument. Through the Infrafrontier MoU national governments have committed to the financial support of the national facilities contributing to the international network of research infrastructures. Infrafrontier functions therefore, like ELIXIR, mainly as a coordinating project, with the national partners contributing funds in a flexible way to allocation of national capacities to the Infrafrontier Research Infrastructure, and provides value through coordination, reducing duplication and pooling of expertise. It remains to be



seen whether and how widely strategies such as these can be implemented outside Europe, given the heterogeneity and relative inflexibility of national science policy contexts, funding priorities, and legal environments.

Funders emphasized that archives need to shoulder the primary responsibility for their own success and financial viability. However, archive managers countered that even when archives become self-sustaining for daily operations (e.g. the KOMP Repository), external public sector funding remains crucial for infrastructure (e.g. equipment and estates), services, and long term sustainability. Moreover, the plant archive TAIR experienced funding cuts even with a record of high achievements [63]. TAIR consequently resorted to charging for-profit, non-profit, and academic users for access. Even with such a subscription model, TAIR can no longer undertake some activities that added value to its data [68–70]. Moreover, adoption of subscription to raise revenues has drawbacks [39]. Archive personnel have the onerous task of monitoring users and ensuring payments. Grant-funded researchers, internationally facing budget constraints, may see subscription-based access as a disincentive. Subscription lowers data integration as paid-access databases may not share their data with free databases. Tiered subscription models, with basic data sets openly accessible and additional charges for enhanced data, tools, and services, are neither tenable nor equitable. In addition, researchers may not provide data to subscription-based databases if their funder stipulates open-access publication. Non-contribution of new data would depreciate quality and usage of databases over time, leading perhaps to their closure.

With new disruptive technologies, such as gene editing, challenging the utility of ES cells in mouse research, funders need to consider the long-term strategic value of their support to repositories, which will need to adapt to new technologies to retain user bases and, more importantly, to disseminate novel but also standardized, reproducible, and affordable reagents to the wider community. It must be remembered that such public repositories serve an important role for industry as well as for academic and non-profit researchers. Moreover, the aims of public infrastructures are distinct from those of purely commercial operations precisely because they can hedge public investment in research against novel uses and developments in enabling technologies.

Regarding income-generation models that archives can adopt, key revenue pipelines can be opened by attracting industry users, who may be asked to pay higher costs than their academic counterparts. However, the KOMP and EUComm experiences indicate that for distribution to the commercial sector, resource developers and distributing archives need external protections against rent-seeking litigation by intellectual property rights (IPR) holders, who may enforce claims on components of materials, methods, and technologies used at various points in resource development [10,11].

### *Governance of archives to support community-level research activities*

The governance of a research commons needs to ensure that research tools are made available and accessible via archives [7,9]. Archive use for deposits and withdrawals, coupled to public attribution in peer-reviewed literature, facilitates accessibility of reagents and data, non-duplicative research, experimental testing

of existing knowledge, and generation of new insights [1,2,8]. The return contribution of secondary and tertiary data and materials by end users enhances the value of the original resources and facilitates the iterative process of knowledge testing and generation. Our data indicate a need for funders, archive managers, and publishers to support research infrastructures via a coordinated tripartite approach that (1) develops incentives for deposit of data and materials into archives, (2) monitors and attributes deposits, and (3) judiciously enforces policies for data and materials sharing [8,59,71–73].

Compared to archive managers, funders and journal publishers have the strategic positioning and authority to jointly translate sharing agreements into community practice. An issue for further investigation is the feasibility and acceptability of digital governance mechanisms to integrate the actions of funders, publishers, and researchers to ensure sharing via the use of public archives. For example, publishers may collate accession numbers for data and material deposits, which could be made available to funders to inform their reviews of project output and applications for grant renewals. Simultaneously, peer reviewers of funding applications could receive guidance from funders on how to assess such records in a manner similar to assessment of peer-reviewed publications. The expansion of formats and norms of publication and citation could encourage researchers to share metadata, which are valuable for experimental testing of published results. Researchers could also be motivated to publicly deposit their data and materials if institutions viewed related records as markers of productivity.

In addition, archives require attribution to build their visibility to attract users and to measure and demonstrate their utility to prospective funders. Journals may be instrumental in ensuring attribution that points to published research reagents in archives. Finally, following Ostrom and Hess [7] on ensuring compliance with the rules of using the commons, funders need to appropriately enforce sharing policies and send an unequivocal message that publicly funded research reagents need to be made speedily and widely accessible. Enforcement is also important to curb free riders, or users who extract resources from the commons without contributing anything in return. For example, a concern, particularly of individual researchers, is competitors withdrawing deposited strains and wresting scientific priority from the original depositors; or industry using archived mouse models without making adequate compensation. In the current scenario, however, only a few powerful and influential funders (e.g. the NIH and the Wellcome Trust) are ready to enforce compliance with sharing policy. Any prospective transnational oversight of post-project sharing of materials and data will need consistent action by the relevant gatekeepers. While such actions are not yet very evident, there is some recognition of their necessity.

### **Conclusion**

Our analysis indicates an urgent need for the design and implementation of new policies to support community archives that serve both large-scale, transnational resource projects and individual research groups. While high-impact resources can be built through strong leadership and project management, their long-term accessibility and sustainability require supporting reforms in funding and governance, implemented via the coordinated action of archive managers, funders, journals, and research institutions.

## Acknowledgments

This work was supported by the NorCOMM2 Project funded by Genome Canada [AM and TB]; the Ontario Genome Institute [TB]; and the Canadian Stem Cell Network [TB]. The authors have no competing interests. The funders of the study exerted no influence on the design and conduct of the study or on the analysis and presentation of results. We thank Lesley Dacks, Lorna Skaley and Dr Ann Flenniken for research support and project coordination.

## References

- [1] Schofield PN, Eppig J, Huala E, de Angelis MH, Harvey M, Davidson D, et al. Research funding. Sustaining the data and bioresource commons. *Science* 2010;330(October (6004)):592–3. <http://dx.doi.org/10.1126/science.1191506>.
- [2] Bubela T, FitzGerald GA, Gold ER. Recalibrating intellectual property rights to enhance translational research collaborations. *Sci Transl Med* 2012;4(February (122)):122cm3. <http://dx.doi.org/10.1126/scitranslmed.3003490>.
- [3] National Institutes of Health [Internet]. Final NIH Statement on Sharing Research Data. Available from: <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html> [published online 26.02.03; cited 26.12.14].
- [4] BioSharing [Internet]. BioSharing policies: a catalogue of data preservation, management and sharing policies from international funding agencies and regulators. Available from: <http://biosharing.org/policies> [updated 2014; cited 26.12.14].
- [5] Bloom T, Ganley E, Winker M. Data access for the open access literature: PLOS's data policy. *PLoS Biol* 2014;12(February):e1001797. <http://dx.doi.org/10.1371/journal.pbio.1001797>.
- [6] PLOS ONE [Internet]. PLOS Editorial and Publishing Policies: Sharing of Data, Materials, and Software. Available from: <http://www.plosone.org/static/policies#sharing> [published 2014; cited 26.12.14].
- [7] Ostrom E, Hess C. A framework for analyzing the knowledge commons. In: Hess C, Ostrom E, editors. *Understanding knowledge as a commons: from theory to practice*. Cambridge, MA: MIT Press; 2007. p. 41–82.
- [8] Schofield PN, Bubela T, Weaver T, Portilla L, Brown SD, Hancock JM, et al. Post-publication sharing of data and tools. *Nature* 2009;461(September (7261)):171–3. <http://dx.doi.org/10.1038/461171a>.
- [9] Dedeurwaerdere T. Global microbial commons: institutional challenges for the global exchange and distribution of microorganisms in the life sciences. *Res Microbiol* 2010;161(July–August (6)):414–21. <http://dx.doi.org/10.1016/j.resmic.2010.04.012>.
- [10] Bubela T, Schofield PN, Ryan CD, Adams R, Einhorn D. Managing intellectual property to promote pre-competitive research: the mouse as a model for constructing a robust research commons. *J Law Inf Sci* 2012;22(October (1)):EAP1–24.
- [11] Mishra A, Bubela T. Legal agreements and the governance of research commons: lessons from materials sharing in mouse genomics. *OMICS* 2014;18(April (4)):254–73. <http://dx.doi.org/10.1089/omi.2013.0158>.
- [12] Ankeny RA, Leonelli S. What's so special about model organisms? *Stud Hist Philos Sci A* 2011;42(June (2)):313–23. <http://dx.doi.org/10.1016/j.shpsa.2010.11.039>.
- [13] Jackson Laboratory [Internet]. History of the Mouse Mutant Resource. Available from: <http://mousemutant.jax.org/MMR-history.html> [updated 2014; cited 26.12.14].
- [14] Kaye J. From single biobanks to international networks: developing e-governance. *Hum Genet* 2011;130(September (3)):377–82. <http://dx.doi.org/10.1007/s00439-011-1063-0>.
- [15] Meijer I, Molas-Gallart J, Mattsson P. Networked research infrastructures and their governance: the case of biobanking. *Sci Public Policy* 2012;39(May (4)):491–9. <http://dx.doi.org/10.1093/scipol/scs033>.
- [16] Crosswell LC, Thornton JM. ELIXIR: a distributed infrastructure for European biological data. *Trends Biotechnol* 2012;30(May (5)):241–2. <http://dx.doi.org/10.1016/j.tibtech.2012.02.002>.
- [17] ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489(September (7414)):57–74. <http://dx.doi.org/10.1038/nature.11247>.
- [18] Brown SD, Moore MW. Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium. *Dis Model Mech* 2012;5(May (3)):289–92. <http://dx.doi.org/10.1242/dmm.009878>.
- [19] Chandras C, Weaver T, Zouberakis M, Smedley D, Schughart K, Rosenthal N, et al. Models for financial sustainability of biological databases and resources. *Database (Oxf)* 2009;2009:bap017. <http://dx.doi.org/10.1093/database/bap017>.
- [20] Stenson PD, Ball EV, Mort M, Phillips AD, Shaw K, Cooper DN. The Human Gene Mutation Database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution. *Curr Protoc Bioinform* 2012, September. <http://dx.doi.org/10.1002/0471250953.bi0113s39>. Chapter 1, Unit 1.13.
- [21] Baker M. Repositories share key research tools. *Nature* 2014;505(January (7483)):272. <http://dx.doi.org/10.1038/505272a>.
- [22] Singh P, Schimenti JC, Bolcun-Filas E. A mouse geneticist's practical guide to CRISPR applications. *Genetics* 2014;199(September (1)):1–15. <http://dx.doi.org/10.1534/genetics.114.169771>.
- [23] Bradley A, Anastassiadis K, Ayadi A, Battey JF, Bell C, Birling MC, et al. The mammalian gene function resource: the International Knockout Mouse Consortium. *Mamm Genome* 2012;23(October (9–10)):580–6. <http://dx.doi.org/10.1007/s00335-012-9422-2>.
- [24] Austin CP, Battey JF, Bradley A, Bucan M, Capecchi M, Collins FS, et al. The knockout mouse project. *Nat Genet* 2004;36(September (9)):921–4. <http://dx.doi.org/10.1038/ng0904-921>.
- [25] Andrews TD, Whittle B, Field MA, Balakrishnan B, Zhang Y, Shao Y, et al. Massively parallel sequencing of the mouse exome to accurately identify rare, induced mutations: an immediate source for thousands of new mouse models. *Open Biol* 2012;2(May (5)):120061. <http://dx.doi.org/10.1098/rsob.120061>.
- [26] Collins FS, Rossant J, Wurst W. A mouse for all reasons. *Cell* 2007;128(January (1)):9–13. <http://dx.doi.org/10.1016/j.cell.2006.12.018>.
- [27] Auwerx J, Avner P, Baldock R, Ballabio A, Balling R, Barbacid M, et al. The European dimension for the mouse genome mutagenesis program. *Nat Genet* 2004;36(September (9)):925–7. <http://dx.doi.org/10.1038/ng0904-925>.
- [28] Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, Iyer V, et al. A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* 2011;474(June (7351)):337–42. <http://dx.doi.org/10.1038/nature.10163>.
- [29] Beckers J, Wurst W, de Angelis MH. Towards better mouse models: enhanced genotypes, systemic phenotyping and envirotype modelling. *Nat Rev Genet* 2009;10(June (6)):371–80. <http://dx.doi.org/10.1038/nrg.2578>.
- [30] Brown SD, Hancock JM, Gates H. Understanding mammalian genetic systems: the challenge of phenotyping in the mouse. *PLoS Genet* 2006;2(August (8)):e118. <http://dx.doi.org/10.1371/journal.pgen.0020118>.
- [31] Gates H, Mallon AM, Brown SD, EUMODIC Consortium. High-throughput mouse phenotyping. *Methods* 2011;53(April (4)):394–404. <http://dx.doi.org/10.1016/j.ymeth.2010.12.017>.
- [32] International Mouse Phenotyping Consortium [Internet]. Concise Business Plan. Available from: [http://www.mousephenotype.org/sites/beta.mousephenotype.org/files/mousephenotype\\_files/IMPC\\_Business\\_Plan.pdf](http://www.mousephenotype.org/sites/beta.mousephenotype.org/files/mousephenotype_files/IMPC_Business_Plan.pdf) [published online 15.11.10; cited 26.12.14].
- [33] Fuchs H, Gailus-Durner V, Adler T, Aguilar-Pimentel JA, Becker L, Calzada-Wack J, et al. Mouse phenotyping. *Methods* 2011;53(February (2)):120–35. <http://dx.doi.org/10.1016/j.ymeth.2010.08.006>.
- [34] Kollmus H, Post R, Brielmeier M, Fernández J, Fuchs H, McKelvie C, et al. Structural and functional concepts in current mouse phenotyping and archiving facilities. *J Am Assoc Lab Anim Sci* 2012;51(July (4)):418–35. PMID: PMC3400690.
- [35] Koscielny G, Yaikhom G, Iyer V, Meehan TF, Morgan H, Atienza-Herrero J, et al. The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. *Nucleic Acids Res* 2014;42(January (Database issue)):D802–9. <http://dx.doi.org/10.1093/nar/gkt977>.
- [36] International Mouse Phenotyping Consortium [Internet]. IMPC Governance and Coordination. Available from: <https://www.mousephenotype.org/sites/mousephenotype.org/files/IMPC%20Governance%20and%20Coordination%20v11%20June%202013.pdf> [published online 11.06.13; cited 26.12.14].
- [37] National Institutes of Health [Internet]. NIH to make a mightier mouse resource for understanding disease. Available from: <http://www.nih.gov/news/health/sep2011/nhgri-29.htm> [published online 29.09.11; cited 26.12.14].
- [38] Corbin J, Strauss A. *Basics of qualitative research: grounded theory procedures and techniques*, 3rd ed., Wiltshire, UK: Sage; 2008.
- [39] Bastow R, Leonelli S. Sustainable digital infrastructure. Although databases and other online resources have become a central tool for biological research, their long-term support and maintenance is far from secure. *EMBO Rep* 2010;11(October (10)):730–4. <http://dx.doi.org/10.1038/embor.2010.145>.
- [40] INFRAFRONTIER Consortium. INFRAFRONTIER-providing mutant mouse resources as research tools for the international scientific community. *Nucleic Acids Res* 2014. pii:gku1193.
- [41] European Commission [Internet]. Legal framework for a European Research Infrastructure Consortium – ERIC Practical Guidelines. Available from: [http://ec.europa.eu/research/infrastructures/pdf/eric\\_en.pdf](http://ec.europa.eu/research/infrastructures/pdf/eric_en.pdf) [published online 2010 April; cited 26.12.14].
- [42] UniProt Consortium. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res* 2010;38(January (Database issue)):D142–8. <http://dx.doi.org/10.1093/nar/gkp846>.

- [43] US Government (General Services Administration, Department Of Defense, National Aeronautics and Space Administration) [Internet]. Title 48: Federal Acquisition Regulations System, Part 52—Solicitation Provisions and Contract Clauses, Subpart 52.2—Text of Provisions and Clauses, 52.227-1 Authorization and Consent. Available from: <https://acquisition.gov/far/current/pdf/FAR.pdf> [published online 2005 March; cited 26.12.14].
- [44] genOway [Internet]. genOway offers industry scientists immediate access to EUComm conditional KO mouse models. Lyon, France: genOway. Available from: <http://www.genoway.com/data/fichiers/file/contenu/products-services/eucomm/eucomm-genoway-agreement> [published online 02.07.14; cited 26.12.14].
- [45] Check, E. [Internet]. NIH intervenes in Alzheimer's mouse lawsuit. Available from: [http://blogs.nature.com/news/2011/06/nih\\_intervenes\\_in\\_alzheimers\\_m.html](http://blogs.nature.com/news/2011/06/nih_intervenes_in_alzheimers_m.html) [published online 17.06.11; cited 26.12.14].
- [46] Hayden EC. Patent dispute threatens US Alzheimer's research. *Nature* 2011;472(April (7341)):20. <http://dx.doi.org/10.1038/472020a>.
- [47] United States Court of Appeals for the Federal Circuit [Internet]. AIA vs Avid Radiopharmaceuticals and The Trustees of the University of Pennsylvania and University of South Florida Board of Trustees. Available from: <http://docs.justia.com/cases/federal/appellate-courts/cafc/13-1544/13-1544-2014-05-16.pdf> [filed 2014 16.05.14; cited 26.12.14].
- [48] Bull KR, Rimmer AJ, Siggis OM, Miosge LA, Roots CM, Enders A, et al. Unlocking the bottleneck in forward genetics using whole-genome sequencing and identity by descent to isolate causative mutations. *PLoS Genet* 2013;9(January (1)):e1003219. <http://dx.doi.org/10.1371/journal.pgen.1003219>.
- [49] Gaj T, Gersbach CA, Barbas 3rd CF. ZFN-TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* 2013;31(July (7)):397–405. <http://dx.doi.org/10.1016/j.tibtech.2013.04.004>.
- [50] Carstea AC, Pirity MK, Dinnyes A. Germline competence of mouse ES and iPS cell lines: chimera technologies and genetic background. *World J Stem Cells* 2009;1(December (1)):22–9. <http://dx.doi.org/10.4252/wjsc.v1.i1.22>.
- [51] Lin Y, Cradick TJ, Brown MT, Deshmukh H, Ranjan P, Sarode N, et al. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res* 2014;42(July (11)):7473–85. <http://dx.doi.org/10.1093/nar/gku402>.
- [52] Nature Editorial [No authors listed]. Still much to learn about mice. *Nature* 2014;509(May (7501)):399.
- [53] Halberg RB, Chen X, Amos-Landgraf JM, White A, Rasmussen K, Clipson L, et al. The pleiotropic phenotype of Apc mutations in the mouse: allele specificity and effects of the genetic background. *Genetics* 2008;180(September (1)):601–9. <http://dx.doi.org/10.1534/genetics.108.091967>.
- [54] Russell W, Burch R [Internet]. The Principles of Humane Experimental Technique. Maryland, US: Johns Hopkins University. Available from: [http://altweb.jhsph.edu/pubs/books/humane\\_exp/het-toc](http://altweb.jhsph.edu/pubs/books/humane_exp/het-toc) [updated 2014; 26.12.14].
- [55] Contreras JL. Information access. Prepublication data release, latency, and genome commons. *Science* 2010;329(July (5990)):393–4. <http://dx.doi.org/10.1126/science.1189253>.
- [56] Leonelli S. When humans are the exception: cross-species databases at the interface of biological and clinical research. *Soc Stud Sci* 2012;42(April (2)):214–36.
- [57] Bubela T, Guebert J, Mishra A. Use and misuse of material transfer agreements: lessons in proportionality from research repositories and litigation. *PLOS Biol* 2015;13(February (2)):e1002060. <http://dx.doi.org/10.1371/journal.pbio.1002060>.
- [58] Leonelli S, Ankeny RA. Re-thinking organisms: the impact of databases on model organism biology. *Stud Hist Philos Biol Biomed Sci* 2012;43(March (1)):29–36. <http://dx.doi.org/10.1016/j.shpsc.2011.10.003>.
- [59] Birney E, Hudson TJ, Green ED, Gunter C, Eddy S, Rogers J, et al. Prepublication data sharing. *Nature* 2009;461(September (7261)):168–70. <http://dx.doi.org/10.1038/461168a>.
- [60] McEntyre J, Ostell J, editors [Internet]. Glossary: The NCBI Handbook. Bethesda, Maryland, National Center for Biotechnology Information. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK21101/> [published 2002; cited 26.12.14].
- [61] Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res* 2014. <http://dx.doi.org/10.1093/nar/gku1216>.
- [62] Van Noorden R. Funders punish open-access dodgers. *Nature* 2014;508(April (7495)):161. <http://dx.doi.org/10.1038/508161a>.
- [63] Leonelli S. Global data for local science: assessing the scale of data infrastructures in biological and biomedical research. *BioSocieties* 2013;8(4):449–65. <http://dx.doi.org/10.1057/biosoc.201323>.
- [64] ELIXIR [Internet]. ELIXIR Consortium Agreement. Available from: [https://www.elixir-europe.org/documents/elixir\\_consortium\\_agreement](https://www.elixir-europe.org/documents/elixir_consortium_agreement) [published online January 2014; cited 06.08.15].
- [65] ELIXIR [Internet]. Implementation of ELIXIR: Frequently Asked Questions on Legal and Governance Issues. Available from: [https://www.elixir-europe.org/system/files/faq\\_on\\_elixirs\\_legal\\_framework\\_january\\_2014.pdf](https://www.elixir-europe.org/system/files/faq_on_elixirs_legal_framework_january_2014.pdf) [published online January 2014; 06.08.15].
- [66] ELIXIR [Internet]. ELIXIR hub financial plan (2014–2018). Available from: [https://www.elixir-europe.org/system/files/elixir\\_hub\\_financial\\_plan\\_2014-2018.pdf](https://www.elixir-europe.org/system/files/elixir_hub_financial_plan_2014-2018.pdf) [published online January 2014; 06.08.15].
- [67] Attwood TK, Agit B, Ellis LBM. Longevity of biological databases. *EMBnet J* 2015;21(May):e803. <http://dx.doi.org/10.14806/ej.21.0.803>.
- [68] Abbott A. Plant genetics database at risk as funds run dry. *Nature* 2009;462(November (7271)):258–9. <http://dx.doi.org/10.1038/462258b>.
- [69] International Arabidopsis Informatics Consortium. An international bioinformatics infrastructure to underpin the Arabidopsis community. *Plant Cell* 2010;22(August (8)):2530–6. <http://dx.doi.org/10.1105/tpc.110.078519>.
- [70] Hayden EC [Internet]. Popular plant database set to charge users. Available from: <http://www.nature.com/news/popular-plant-database-set-to-charge-users-1.13642> [published online 31.08.13; 26.12.14].
- [71] Field D, Sansone SA, Collis A, Booth T, Dukes P, Gregurick SK, et al. Megascience. 'Omics data sharing. *Science* 2009;326(October (5950)):234–6. <http://dx.doi.org/10.1126/science.1180598>.
- [72] Chervitz SA, Deutsch EW, Field D, Parkinson H, Quackenbush J, Rocca-Serra P, et al. Data standards for Omics data: the basis of data sharing and reuse. *Methods Mol Biol* 2011;719(January):31–69. [http://dx.doi.org/10.1007/978-1-61779-027-0\\_2](http://dx.doi.org/10.1007/978-1-61779-027-0_2).
- [73] Expert Advisory Group On Data Access [Internet]. Establishing incentives and changing cultures to support data access. Available from: [http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh\\_peda/documents/web\\_document/wtp056495.pdf](http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh_peda/documents/web_document/wtp056495.pdf) [published May 2014; cited 26.12.14].