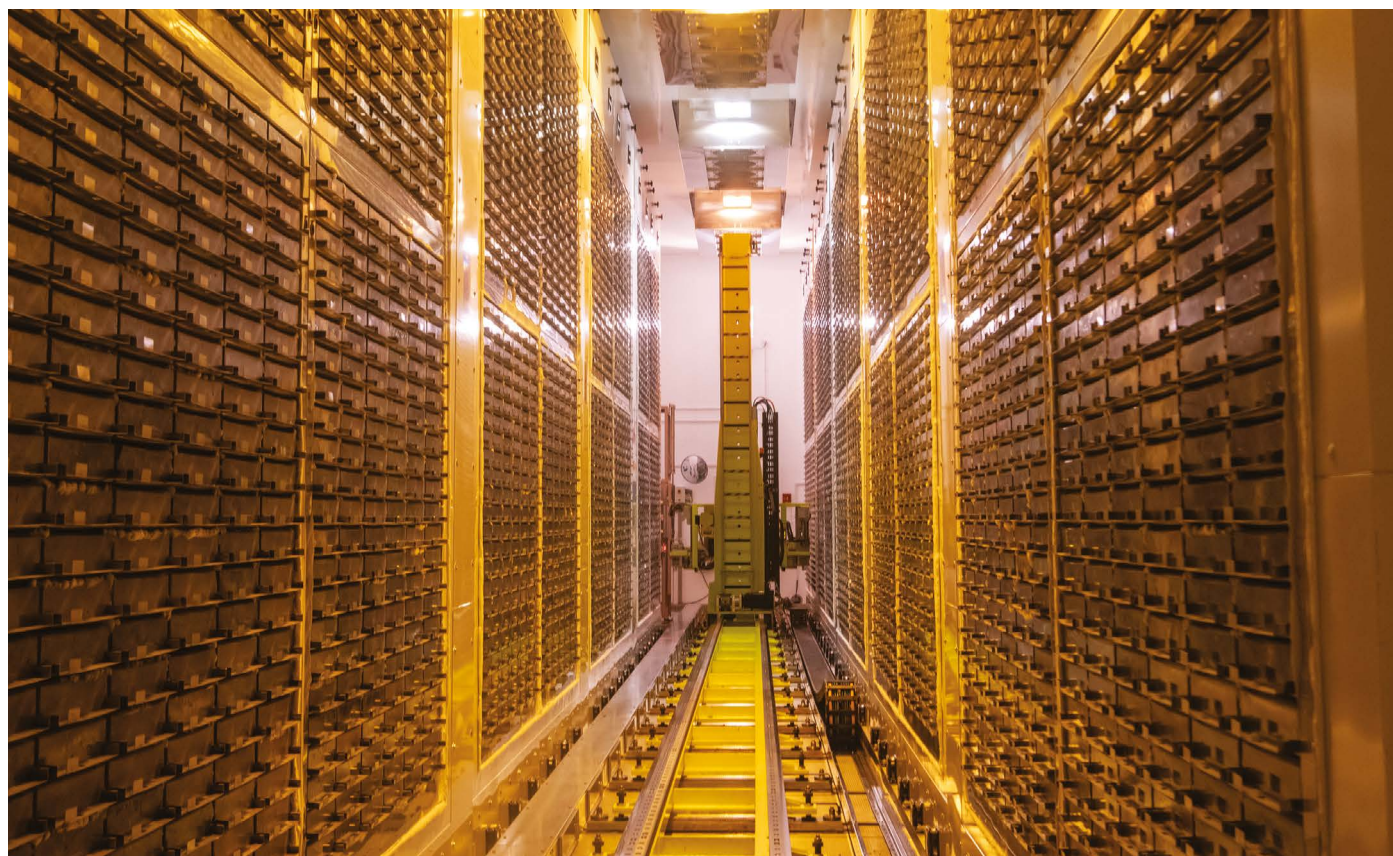




DAVID GUTTRIDGE/UK BIOBANK

A freezer at the UK Biobank stores some of the repository's ten million biological samples of blood, saliva and urine at -80°C .

USING BIOBANKS TO BOOST RESEARCH: A HOW-TO GUIDE

From large national databases to bespoke sample collections, biobanks offer a wealth of avenues for scientific enquiry. **By Jyoti Madhusoodanan**

Rafaella Rogatto De Faria was nearing the end of her PhD when her adviser proposed a fresh project. The idea was to analyse genetic, imaging and surgical-outcome data, to find biomarkers that could help to identify which people with osteoarthritis would respond best to knee-replacement surgery. De Faria, an athlete and a biomedical engineer at the University of São Paulo, Brazil, knew the profound impact of cartilage and joint injuries on people's lives, and stayed on to pursue the project after she defended her PhD in July 2024.

She and her colleagues began by gathering data from people being treated for osteoarthritis at the university. The team has a cohort of 200 individuals so far, with data gathered over the past two years. "We are actually creating our own biobank," De Faria says. "We don't have this yet here."

But the team also wanted to validate its results

against a larger data set. Colleagues suggested looking to the UK Biobank, a collection of images, clinical and genetic data and physical samples from 500,000 individuals in the United Kingdom, some of whom have been studied since 2006. Within a few months of applying for access, De Faria had at her fingertips data from 40,000 people with osteoarthritis, half of whom had undergone total-knee-replacement surgeries. That's a 200-fold increase over her original cohort. "The data were exactly what we were needing," De Faria says.

De Faria's data needs are not unique. Around the world, researchers studying human health often find themselves in need of more, and more-diverse, samples. They could try to collect them themselves or track down existing samples by contacting researchers in the same field of study. Alternatively, they could reach out to entities that have been specifically created to share these resources: biobanks.

Biobanks range in size from the scale of a single laboratory, such as the collection that De Faria and her adviser started, to that of the UK Biobank, one of the world's largest such collections. Most contain both data and physical samples that researchers can request for their studies. And although some smaller banks aim to serve researchers at a single institution, large-scale initiatives such as the UK Biobank, the Mexico City Prospective Study and the All of Us initiative by the US National Institutes of Health (NIH) are designed to meet the needs of the global research community.

For many researchers, biobanks can mean the difference between a successful project and one that stalls for want of crucial data. Yet most biobanks are underutilized, with some surveys suggesting that less than 10% of banked samples get used, says health-policy researcher Amanda Rush at the University of Sydney in Australia.

In planning a biobank-enhanced project, researchers must weigh the pros and cons of creating their own collection of samples and data against the support that larger, existing biobanks can offer. They must also factor in practical considerations such as cost, data security and the legalities of shipping biological materials, Rush says.

And then there are the more strategic considerations. Some projects might be best served by having a large number of samples, but others might benefit from a bespoke collection that offers richer metadata for each specimen, Rush says. There are different scenarios for which each of these biobanks “comes to the fore”, she says.

Roots of biobanks

Biobanking as an enterprise stems from technological advances that began in the late 1980s, says Peter Watson, who leads biobanking services at BC Cancer Research, a research institute in Vancouver, Canada. Speedier DNA-sequencing technologies, faster computing and larger, more powerful databases meant that biological data could be collected and reused endlessly. But efforts to create repositories of data and samples were mostly siloed and ad hoc. “It was just sort of individual efforts in different institutions,” he says.

As a graduate student studying rare paediatric tumours in the early 1990s, Jennifer Byrne and her adviser relied heavily on tissues that had been surgically removed. The specimens were often large – on one occasion, Byrne remembers rushing to the hospital to receive a grapefruit-sized sample – and patients donated them in the hope that others with the same disease would benefit. “There were no cell lines for those cancer types, so we had to study human material,” says Byrne, who is now a molecular oncologist at the University of Sydney. The result was effectively a biobank, “but we didn’t even really realize that we were doing that”.

Although samples gathered in this way are not freely available, because of issues around consent, researchers who can demonstrate funding and ethical approval can approach the custodians for collaborations, Byrne says.

This pathway for external investigators to tap into a collection is what separates a biobank from a stash in a lab freezer, Byrne says. “Biobanks are designed to be reused for different purposes, by different people.”

Understanding the access policies early on is crucial to success, she adds. “Do they provide samples to anybody, or are they largely set up to serve the needs of a single network of researchers?” Finding smaller biobanks, with more-restricted access, can be tricky because they are generally not well advertised. In 2016, Byrne and her colleagues created a biobank registry for the Australian state of New South Wales, allowing researchers



A UK Biobank researcher handles frozen samples.

to find information about resources in their area and register their own biobanks (<https://nsw.biobanking.org>).

The Royal College of Surgeons in Ireland (RCSI) has taken a similar approach, curating several disease-specific collections created by individual RCSI researchers into an institutional biobanking service at its headquarters in Dublin. The biobank streamlines the process of donor consent for samples to be used in research. Researchers who wish to contribute samples can ask to have their own collections added, and would-be collaborators can apply to use the materials or data. But the clinicians who gathered the samples remain closely involved in their use.

Indeed, researchers wanting to access the materials or data should plan on collaborations rather than treating the biobank purely as a vault to extract information, says RCSI geneticist Gianpiero Cavalleri. “We want it to be used as much as possible,” Cavalleri says. “But the typical access model is in collaboration with the investigator.”

Scaling up

Despite the intent to share, tapping into a restricted-access biobank can be challenging for many researchers, Rush says. Shipping can be up to US\$50 per sample, meaning it could cost thousands of dollars to acquire enough material for a large study. Furthermore, the complex legal and other agreements required to transport biological samples or share data securely can stymie early-career researchers or those without large pots of funding. Turning to larger biobanks can help to surmount these barriers, because they might have systems in place to help with the logistics.

Another consideration, says clinical researcher Alex Chaitoff at the University of Michigan in Ann Arbor, is the breadth of

samples available. For his work, Chaitoff often uses large databases, such as the NIH All of Us biobank and the US Centers for Disease Control and Prevention’s National Health and Nutrition Examination Survey (NHANES). This is the only national US health database that includes health and nutrition information for people of all ages, with roughly 5,000 participants added each year. These data sources “are much more likely to be nationally representative”, Chaitoff says. All of Us is especially valuable, he adds, because it includes groups that have been historically excluded from scientific research, such as Native American communities (The All of Us Research Program Genomics Investigators. *Nature* **627**, 340–346; 2024). “They oversample populations that are generally undersampled in research,” he says.

Gaining access

Once researchers find a biobank with the data that they need, they must navigate issues around access – and payment.

Duniel Delgado Castillo, a biomedical engineer at the National Autonomous University of Mexico in Mexico City, was combing through the research literature on physical changes in the brains of people with long COVID when he stumbled across a trove of brain images that he desperately needed. Castillo had already tried, with little success, to reach out to the authors of various studies to access their image collections. By comparison, the resource he discovered, part of the UK Biobank, was much larger, easier to access and seemed to have exactly what he was looking for.

He applied for access early in 2024. Although the initial application suggested there would be a fee in the £3,000–£9,000 (US\$4,000–\$12,000) range for three years of access, the biobank offered him a grant to cover the costs. Within months, he was

working with magnetic resonance imaging (MRI) scans from 1,000 participants, half of whom had long COVID; the other half were matched controls. “If I didn’t have those images, it would be impossible to continue with my investigation,” Castillo says.

Data from the UK Biobank can be used only within its own secure cloud-computing service, the Research Analysis Platform. It includes tools to analyse genomic and translational data, perform statistical analyses on images and other data types, as well as providing machine-learning tools that can be accessed using JupyterLab, an open-source data-science system.

Training to use the biobank’s secure computing space was easy, Castillo and De Faria say. And trying to recreate it on their institutional systems would have been cumbersome and expensive, Castillo adds. “It was a relief for me because all the security of the data is built into the Research Analysis Platform,” he says. “I don’t have to worry about it.”

Before April 2024, researchers from low- and middle-income countries (LMICs) paid a reduced rate, says Lauren Carson, who leads research development for the UK Biobank in London. But the fee was still a deterrent for many. So last year, the organization launched the Global Researcher Access Fund, which waives all fees for researchers in many LMICs.

For now, researchers who are covered by this fund can access only data and not the biobank’s massive repository of frozen blood, urine and saliva samples. The samples are available to researchers not covered by the global fund, but those investigators can’t select or request subsets that meet certain study criteria; instead, they can work either with the entire sample set or with a random selection (costing £5–10 per sample). “We strongly discourage cherry-picking,” Carson says. “Otherwise, you

end up with certain participants where lots and lots of samples have been taken and others where that isn’t the case.”

Carson encourages researchers who are interested in obtaining physical samples to look through guidance on the website first (www.ukbiobank.ac.uk). Chaitoff points to webinars and guidance on the NIH All of Us website (<https://allofus.nih.gov>) – and, if all else fails, he suggests reaching out to that biobank directly. “I have found it very easy to connect with and ask questions of the folks who are overseeing it,” he says.

Because the number and quantity of samples is finite, the UK Biobank emphasizes that research studies should aim to convert samples

“Even with a limited budget you get access to all that data.”

into data that can be reused, effectively turning a limited resource into a limitless one.

Researchers can return unused samples. And, after about nine months of exclusivity, any fresh data must be sent to the biobank, where internal teams track their quality and completeness. “Any data that you generate, we want to be able to pass it on for other researchers to use,” Carson explains.

Seeing beyond specimens

With an increasing number of options for finding and accessing biobanks of all sizes, researchers will need to identify which ones best suit their needs. In a 2022 study, Rush and her colleagues found that half of the researchers surveyed said they had to use resources from more than one biobank for their work (A. Rush *et al.* *Biopreserv. Biobank.* **20**, 271–282; 2022). Failing to find the best fit takes a toll: 60% of respondents said they limited the scope of their research because they couldn’t get the data they wanted.

Cavalleri has used several biobanks for his epilepsy research. When working closely with the relatively small RCSI biobank, researchers can glean valuable information from clinicians about different courses of treatment, people’s responses, the results of ongoing imaging tests and other details that can inform the analysis, Cavalleri adds. “You’re collecting really deep phenotype data.”

He has also used the UK Biobank, which, within a few months, gave him access to genomic, proteomic and other data on thousands of people with epilepsy. “Even with a limited budget,” covering just your own salary and the access fee, “you get access to all that data,” Cavalleri says.

But researchers should also recognize the biases built into such a database, he adds. The UK Biobank is representative of the UK

population, meaning that participants are mostly of European ancestry. They are also all adults over the age of 40, so the biobank might be less useful to a researcher working on rare diseases that shorten lifespan significantly. “You can’t have a scientific infrastructure which is just a large population-based biobank,” he says. “You do need the disease-specific ones as well.”

Ultimately, the services and insight that biobank staff offer can be more valuable than the samples or data that biobanks hold. A biobank sounds like a place where “you walk up to a secure door, and you ask somebody to give you something from inside it”, Watson says. “But we really should be operating as a service group rather than a static entity.”

Efforts to make biobanks adhere to FAIR principles – meaning that their contents are findable, accessible, interoperable and reproducible – are making inroads towards that more collaborative framework. And researchers who launch their own biobanks can help with these efforts by taking steps toward sustainability, Byrne says. For example, they could ensure that participants agree to samples being reused or shared, or they could offer continuing support to users who want to access the samples.

Approaching biobanks with your specific aims and research needs can be helpful, Byrne says. Staff can help researchers to gauge the feasibility of a project and whether a specific biobank, large or small, will meet their needs. Carson concurs, adding that any researcher wondering about how they might use the UK Biobank’s resources should reach out to discuss their questions. “I think it’s very important for researchers and biobanks to talk together,” Byrne says.

As a user of both institutional and national biobanks, Chaitoff emphasizes the value of such discussion. He has used health data from wearable devices included in the All of Us database, and was hesitant when he approached the research group at Vanderbilt University in Nashville, Tennessee, that was contributing the data. However, group members helped him to work with the information and talked over how he could best use their resources. The discussions gave him the confidence to continue partnering with biobanks and their curators.

Hearing that those who run biobanks aim to help researchers – rather than acting as gatekeepers of resources – was “kind of this awakening for me”, Chaitoff recalls. “No one wants these specimens or people’s information to just sit on digital shelves or in freezers. The more people publishing with it, the more it means that the participants who took their time to be part of a biobank are fulfilling their mission of bettering science.”

Jyoti Madhusoodanan is a science writer in Portland, Oregon.



Amanda Rush at the University of Sydney.