Wellcome Collection

We are a free museum and library that aims to challenge how we all think and feel about health. Through exhibitions, collections, live programming, digital stories and services, broadcast and publishing, we create opportunities for people to think deeply about the connections between science, medicine, life and art.

Our digital strategy is to design and build a free and unrestricted digital space that inspires, engages and challenges people to think about what it means to be healthy and human.

As a part of this challenge, we have been developing a world-class online resource for the history of medicine by digitising a substantial proportion of our holdings and making the content freely available on the web, increasing the number of researchers and the diversity of researchers accessing and using our collections.

Digitisation Progress

As of January 2020, we have digitised over 40 million images from our collections and those of our partners, free for anyone to use.

Archives and records
40 collections
2 million images

Printed works: books, pamphlets, reports
200,000 items
35,500,000 images

Printed works: serials
360 titles/10,000 volumes
2,000,000 images

Visual Materials
17,000 items

Audiovisual
3000 media files

Manuscripts
1,200 items, 200,000 images
**Digitisation 2020 – 2025**

Wellcome Collection is a world leader in digitisation. Over the past two decades we have developed highly skilled staff with the experience to carry out high quality digitisation. There are 14 staff members involved in digitisation and three suppliers (two with staff working on site). Team and task management has evolved to support self-organising, flexible teams with a focus on continuous improvement and professional development.

We have streamlined our processes to enable us to handle higher throughputs and a greater diversity of material. Working with development partners, we have employed a range of systems to automate and manage our digitisation pipelines.

This is the foundation that allows us to sustain our digitisation ambitions to provide ever greater access to the full range of our collections into the future.

In terms of progress, we still have a long way to go. Our closest estimate is that no more than 16% of our holdings has been digitised, with large variances between collections. For example, although we have now digitised nearly all our public domain printed works, we have 90 years' worth of manuscripts, and 120 years of archives digitisation ahead of us. With some collections progress is less about how much of a dent we are making in our physical holdings than it is about the quality and impact of our prioritisation strategy.

**People, culture and process**

Although technology and systems greatly help us optimise our workflows, our success in managing large-scale digitisation relies on the teams that we have built, and the way we treat our digitisation efforts as pipelines rather than specific projects.

We subscribe to the “self-organising” principle of team management. Within a framework of collaborative objective-setting on a quarterly basis plus clarity on boundaries, teams can experiment with different approaches to how they work.

A pipeline consists of an ongoing standardised workstream that accommodates the needs of similar physical formats. Rather than defining our work as a series of projects each requiring specific management and workflow considerations, structuring our work as workstreams helps us make the most efficient use of staff availability, equipment set-ups and contractual relationships. Consistency in tracking and scheduling helps everyone work and communicate more efficiently, while maintaining flexibility in terms of prioritising specific items or collections within any given workflow.

Our team of in-house photographers take on complex, high-specification digitisation. They are currently focused on paintings, manuscripts and object-based digitisation, building on decades of expertise and
knowledge of the collections. For larger-scale projects, we use trusted contractors who mostly work on-site alongside staff, allowing us to benefit from economies of scale.

Digitisation management of all strategic digitisation streams, and our Digitisation Service, is largely centralised within a single team, including contract management, scheduling, and the end-to-end workflow. This team works closely with conservation, photographers, and collections staff who can provide specialist support.

Tools

Digitisation team members plan, schedule, track, ingest and preserve our digital assets using a variety of systems to help manage digitisation at scale. Our in-house software development and product management team work closely with external development partners and experts to develop scalable and robust digital asset management and storage solutions. We are able to leverage open-source software throughout most of our ingest and preservation infrastructure, and much of the work we contribute or commission benefits the wider community.

Over the past 10 years we have increasingly automated our ingest workflows using Goobi, an open source workflow system specifically designed to reduce unnecessary manual labour and improve workflow management for digitisation. Goobi is a database and control centre orchestrating the management of digital assets and metadata as they are prepared and organised for ingest. Goobi also produces metadata that describes the logical structure of digital objects, so we can accurately replicate them online – for example, pagination for books, sequencing of multi volume works, access and licencing information, or metadata describing different parts of a work.

On-site scalability of storage and systems has proven more and more difficult as our digital estate has grown. In response to this pressure, we have just completed work to alleviate this capacity bottleneck by moving our digital assets and systems, including Goobi, from local servers to Amazon Web Services (AWS). Virtually unlimited capacity and point-of-need charging structures in the Cloud means we can handle spikes in computation requirements and increase our archival storage requirement in a very cost-effective way. Other major benefits include a standards-based approach to file fixity (BagIt), development of fully open source Storage Services, replication to multiple suppliers and geographical locations, ease of processing and re-processing content such as format conversion and validation, and increased capacity to extract information to help us manage 56 terabytes’ worth of assets.

In addition to the systems that handle all aspects of the digital assets, we use an online subscription database called Quickbase to manage information about the physical collections we are digitising. This database is easily customised by the team, for the team. We can store and retrieve information related to condition assessments, delivery timings and inventories, location and movement of items undergoing digitisation, and progress of these items through the workflow without resorting to countless Excel spreadsheets. Members of other teams involved in supporting digitisation can access the data too, so there is transparency across all teams involved in the digitisation effort.
Over the next 5 years, we will continue to digitise across these 6 workstreams:

**Printed works:** Once we complete digitisation of public domain collections, we will address the 20th century “black hole” with a focus on in-copyright books published up to 1950. Access will be limited due to legal constraints on communication and reuse of these works - many of which are orphan works. We can preserve them to facilitate access under copyright exceptions or by permission from rights holders, now and into the future. We will also fill gaps in some of our journal and report sequences by working with external content holders.

**Audiovisual:** We have scaled up our operations in cooperation with our off-site digitisation partner to address the imminent risk of obsolescence for magnetic media. We plan to digitise our most at-risk formats by 2023, including thousands of audio and video titles. Since moving our preservation services to the Cloud, we are no longer limited in our ability to cope with a large quantity of higher quality formats. Our vendor can upload 2K files directly to our S3 storage area, and all processing can be done on the Cloud, greatly streamlining our workflow. We use have started to use lossless JPEG2000/MXF as our archival format for video (previously MPEG2). This format can be encoded and decoded with tools available on Amazon web services, ensuring long-term access, and the possibility of migration as needed in future.

**Archives:** We will continue to increase access to our unique and distinctive archival holdings, reflecting our collection strengths and priorities at the rate of 400 or more boxes per year. This includes up to 1,000 boxes containing the Family Planning Association archive and others related to our ongoing theme of reproductive health campaigns and services.

**Visual materials:** We will digitise all our oil paintings and selected other paintings, prints, drawings, and engravings. We incorporate user feedback in our methodology for digitising. Based on such feedback, we are digitising our paintings “in the round” including frames and the backs of each painting. Although this requires extra effort, the process of cleaning, repairing, moving and photographing these paintings is so time-consuming, this is one collection where, knowing what the users would potentially like to see, “do it once, do it right” really applies.

**Manuscripts:** We are already well on our way to completing our recipe and medieval western manuscripts, but we will also bring in other collections from our global manuscript holdings, which cover Europe, Asia and Africa. Manuscripts are digitised to a much higher resolution than books or archives, to ensure that we are able to capture not just the textual information, but also the physicality of the items, and their characteristics as objects. Many of these items are very fragile and cannot withstand much handling. Digitisation at a very high resolution means we will not have to handle them again for a very long time.

**Digitisation services:** We provide a free digitisation service for external and internal clients with capacity of up to 400 items per year. Researchers, educators, media professionals and creators are all able to draw on our services to request images, audio and video files from our historic collections. We view this service as user-demand-led digitisation, as much of what we digitise can be made freely online for others to access – items that may not be digitised any time soon through our long-term strategic workstreams.
Reach and Impact

Use of our digital collections has grown steadily over the years as we have added more content. In 2019, we had over 600,000 digital item views on the Wellcome Collection and Library websites. Our digitised printed collections are also stored on the Internet Archive site, where they have been viewed 3.25 million times in the past year. Our audience is global although the majority comes from the UK, Europe and the USA. Through our Imaging Services provision, we interact directly with internal and external users on a daily basis so we know how varied uses can be. Our collections are mostly used by researchers, while print and broadcast media are regular users of individual images and audiovisual materials for both research and creative purposes.

We are working to raise awareness of our digitised collections to new audiences with image galleries, stories and social media posts on Wellcomecollection.org. Throughout 2020 we will be focusing on user-centred design improvements for search and display interfaces, while increasing API functionality. This will transform people's experiences in using our digital content, opening up more opportunities for a greater variety of uses and users.

More information

| **Beta Collections search page**: You can access most of our digitised content here. This site is under active development, and changes frequently. | https://wellcomecollection.org/works (search and view content) |
| | https://wellcomecollection.org/works/progress (information about the beta site) |
| **Library catalogue search**: All digitised content is still available on our Library website, which is no longer under development and due to be decommissioned at the end of 2019. | https://wellcomelibrary.org/ |
| **Information on our work**: Read more about how we are transforming our work with digital material and technology. | https://stacks.wellcomecollection.org/ |
| **Developers**: Information on our APIs, content management and web development. | https://developers.wellcomecollection.org/ |
| **Access policy**: Our digitisation ambitions precipitated an overhaul of our access policy and procedures. View our policy here. | https://wellcomecollection.org/pages/Wvmu3yAAAIUQ4C7F |