

IT and Technical Services Head Office

Home to the Service Desk, IT Security and departmental support services.

[Home](#) [Members](#) **[Blog](#)** [Jobs](#)



Daniel Gant

26 May 2022

Category: Data Storage

Tags: archive, data storage, file replication, fire

Data archiving through the generations



Around 2008 it became clear that the rapid growth in both the size and complexity of EMBL-EBI's data archives required a bespoke, in-house, software-defined object storage solution to ensure its availability, accessibility and recoverability, and the EMBL-EBI FIRE (File Replication) archive was born.

Initially replicating data across two data centres, extreme growth in data deposition through the 2010s led to the addition of a third data centre, introducing further complexity to FIRE, as well as the need to integrate a tape-based cold data storage solution for disaster recovery purposes. By 2017, FIRE was providing a robust, vendor agnostic, geo-dispersed file replication and disaster recovery solution for the biggest and most important data archives at EMBL-EBI. Shortly after this, a RESTful API was developed, transforming FIRE's accessibility and usability for its global user base, allowing access from any compliant compute platform or storage system.

As EMBL-EBI prepares for increasingly large (and difficult to predict) rates of data deposition, FIRE must be continually adapted and improved, as we learn from the past and prepare for the future.

How does FIRE work?

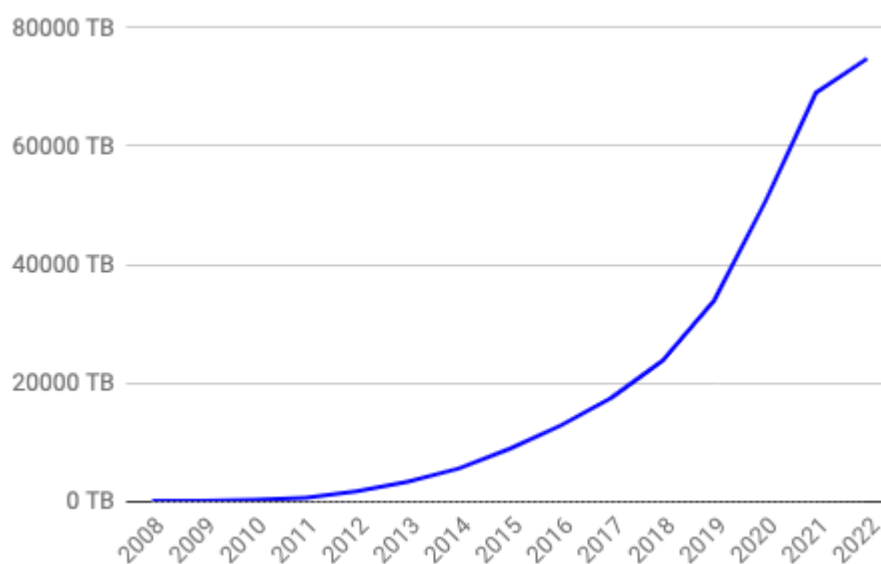
When data is ingested into FIRE, it is placed into one of several storage systems, transparently to the user. A data protection method known as [Erasure coding](#) is used to divide the objects up into sectors in such a way that only a subset is required to

reconstruct the original file or object. These sectors are then distributed across the three data centres, allowing data to be reconstructed from a subset of the data held in any two of the three locations. This means that FIRE may be instantly adapted to cope with the failure of a data centre or a loss of network connectivity. Additionally, a high degree of transparency allows users uninterrupted access to the archive in the event of such a hardware failure.

In addition to resilience, FIRE provides disaster recovery. All data that is written to FIRE is eventually copied to tape storage, which provides a long term, robust, secure and low cost data storage solution, particularly for cold data, to which access is seldom required.

What challenges does FIRE currently face?

During 2020/21, FIRE's growth has been extreme. Traffic resulting from a collaboration between [EGA](#) and [UK Biobank](#), the sudden proliferation of COVID-19 data, and a general increase in data being requested by researchers all over the world almost tipped things over the edge.



Annual growth rates

FIRE currently receives upwards of ten million file requests every day. In order to put its data storage and network bandwidth consumption into perspective, if we liken FIRE's data egress to a movie streaming service such as Netflix, it can be compared to streaming six million movies every month, or just over 200,000 every single day. In terms of ingress, during a busy day, FIRE has to write data at a pace that can be compared to burning 40 DVDs per minute.

What's next for FIRE?

The constant, unrelenting acceleration in data throughput means the technical teams responsible for FIRE at EMBL-EBI are always on their toes. Service & Data Management Coordinator, [Joan Marc Riera Duocastella](#), is realistic about the challenges: "If the data

influx does not slow down, we predict that by 2025 FIRE will have to cope with circa 340PB of unique data, which, after geo-dispersion and disaster recovery requirements, translates to around one exabyte of raw storage...and this should be considered an estimate as current trends point towards more archives wanting to move their long term data to FIRE.” He continues, “we want to keep as much data as possible spinning on disks, so that it is highly available to users, such is the general requirement of our global user community, but it’s not easy! Budget is always an issue, but we have plenty of technical ideas going forward, paying particular attention to the need to ensure FIRE remains built primarily on top of open source technologies, provides standard interfaces, and is maintained with in-house expertise, to ensure we retain control over it and to avoid over reliance on particular vendors or software stacks.”

Marc concludes: “We must understand that, as an archive, FIRE will ultimately outlive us all. Data currently available via a unique file path will still need to be available via that same URL long into the future. The architecture, mechanisms and technologies employed in FIRE will, with certainty, change and evolve over time, likely into something almost unrecognisable by comparison to the current implementation, however access to the archived data must persist indefinitely.”

If you like the idea of working on innovative, complex, open source software solutions such as FIRE, why not check out our latest job vacancies....

<https://www.ebi.ac.uk/careers/jobs>



Categories

[Audio Visual](#) (3)

[Cloud](#) (8)

[Data Storage](#) (3)

[IT Service Management](#) (1)

[IT Support](#) (4)

[Networking](#) (1)

[ResOps](#) (2)

[Search](#) (1)

[TSC Projects](#) (1)

[Web Development](#) (5)

We help scientists exploit complex information to make discoveries that benefit humankind.

SERVICES	RESEARCH	TRAINING	INDUSTRY	ABOUT
Data resources and tools	Publications	Live training	Members Area	Contact us
Data submission	Research groups	On-demand training	Contact	Events
Support and feedback	Postdocs and PhDs	Support for trainers	Industry team	Jobs
Licensing		Contact organisers		News
Long-term data preservation				People and groups
				Intranet for staff

EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK. Tel: +44 (0)1223 49 44 44 Full contact details

Copyright © EMBL 2025 EMBL-EBI is part of the European Molecular Biology Laboratory Terms of use