CRAM

Global Alliance for Genomics & Health

The genomics compression standard

CRAM: community effort

As genomic sequencing increases around the globe, it becomes vital to store this data efficiently and sustainably. GA4GH's CRAM file format for genomic data compression tackles this challenge and helps facilitate global collaboration.

Scroll through the videos below to learn how it has benefited existing users and how you can get involved, too.



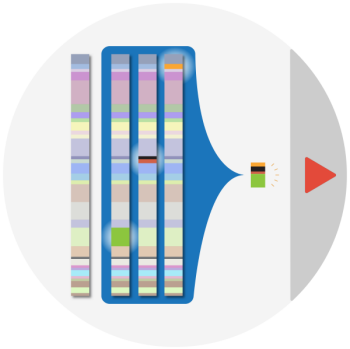


"CRAM is a fundamental part of the GA4GH suite of standards. It's how we think about storing DNA sequences and it works as a package with other standards to allow scientists, healthcare professionals, and commercial researchers to access the information they want when they want it."

Ewan Birney
Chair of GA4GH, Director of EMBL-EBI

About GA4GH

The Global Alliance for Genomics and Health (GA4GH) is an international, nonprofit alliance formed in 2013 to accelerate the potential of research and medicine to advance human health. Bringing together almost 600 leading organisations working in healthcare, research, patient advocacy, life science, and information technology, the GA4GH community is working together to create standards and frameworks to enable responsible and secure global data exchange.





"Even if an organisation is not producing large volumes of data, it can be very beneficial to use the same JIe formats that other organisations are using. The community has a vital role in making sure that a standard is suitable for everybody and not just one individual or one group."

James Bonfield
Wellcome Sanger Institute

CRAM is...



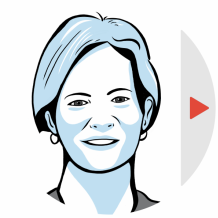
Global
Anthony Philippakis
Broad Institute



Sustainable
Tiffany Boughtwood
Australian Genomics



Economical
Malachi Griffith
VICC



Efficient
Nicola Mulder
H3Africa



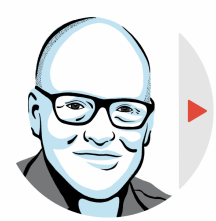
Transformative
Thomas Keane
EMBL-EBI



Open Source
Peter
Counter
Genomics England



Interoperable
Paul
Flicek
Ensembl



Evolving
Albert
Vernon-Smith
University of Michigan



The Future
Pär
Lundin
SciLifeLab



Collaborative
Ira
Hall
McDonnell Genome Institute



"The CRAM file format is essential toward reducing the footprint of genomic data files enabling more efficient, large-scale analyses and queries and also supporting population scale sized projects such as Genomics England. Importantly, the CRAM format was developed by genomics experts within the community to solve a unique challenge in scaling experiments and applications. The adoption of CRAM shows the user benefit when tools are developed by the community for the community."

Susan Tousi
Illumina Inc.

Benefits of CRAM

Reduces disk space and storage costs by 30-50%

Interoperable with other industry standards and best practices

Accurately tracks reference genome, improving integration with the field

Easily transfer and share data with collaborators

Continuous community effort to upgrade the format

Free to the community



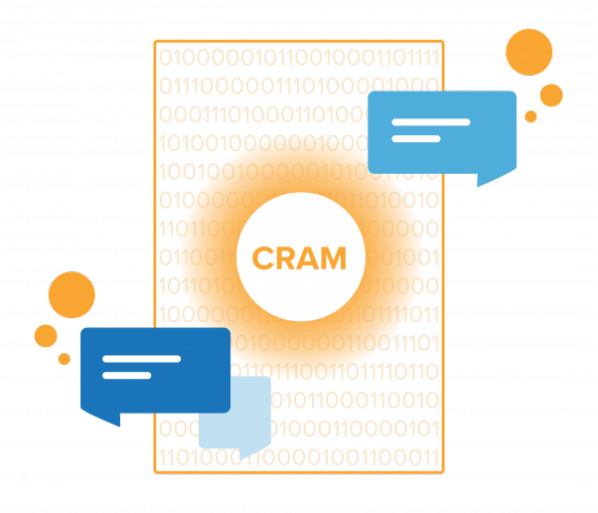
"It becomes absolutely critical that [genomic] data is in a format that can be easily and effectively shared amongst many investigators. In other words, it really becomes very wasteful if an investigator has to go and reprocess or reformat the data every time they get a new dataset."

Stacey Gabriel
Broad Institute

Start the conversation at your organisation

There are many ways to get involved. Start the dialogue at your organisation or institute around using CRAM.

1. Adopt CRAM-enabled tools and libraries (see below)
2. Adopt the CRAM specification
3. Join the CRAM development community
4. Share this page with your colleagues





"Open and encumbered file formats are essential for the science and growing commerce of this industry. The innovations in genomics that will serve humanity are in the interpretation of the data. GA4GH standards allow data and algorithm sharing between institutions, vital for a vibrant informatics ecosystem."

Warren Kaplan
Garvan Institute

Used globally by...

EMBL-EBI · BROAD INSTITUTE · illumina · SciLifeLab · wellcome sanger institute

Australian Genomics · Genomics england · NHLBI TOPMed · H3Africa Human Heredity & Health in Africa · ...and more

Libraries and tools supporting CRAM

Software Libraries

htslib | htsjdk | PySam | Bio::DB::HTS | RustBio

Tools

Samtools | GATK | Picard | IGV | Crumble

Data Archives

European Nucleotide Archive (ENA) | European Genome-phenome Archive (EGA)

Genome Browsers

ENSEMBL | JBrowse | UCSC Genome Browser