

Mandated data types

For the following data types submission to a community-endorsed, public repository is mandatory. Persistent identifiers (DOIs and accession numbers) assigned to the data by the repository must be appropriately cited and referenced in the published article.

Data types	Repository options
DNA sequence data*	Any INSDC member repository / Genome Sequence Archive (GSA)
RNA sequence data*	
Genome assembly data*	
Protein sequence data	Uniprot
Proteomics data	Any ProteomeXchange member repository
Genetic variation data	dbSNP (human variations less than 50bp)
	dbVar (human variations greater than 50bp)
	European Variation Archive (EVA) (all species)
	Genome Sequence Archive for Human (human variation)
Functional genomics data	dbGAP
	ClinVAR
	The European Genome-phenome Archive (EGA)
Macromolecular structure data	Worldwide Protein Data Bank (wwPDB)
	Biological Magnetic Resonance Data Bank (BMRB)
	Electron Microscopy Data Bank (EMDB)
Gene expression data	Gene Expression Omnibus (GEO)
	ArrayExpress
Crystallographic data for small molecules	Cambridge Structural Database (CSD/CCDC)
	Crystallography Open Database (COD)

*Novel DNA sequence, novel RNA sequence, and novel genome assembly data must be deposited to repositories that are part of the [International Nucleotide Sequence Collaboration](#) (INSDC), or those which are working towards INSDC inclusion (as listed in the table), unless there are privacy or ethics restrictions that prevent open sharing of such data. Novel DNA sequence, novel RNA sequence, and novel genome assembly data may in addition be deposited to any other repository (including regional or national repositories) as required.