

FAIRness assessment for UniProt datasets

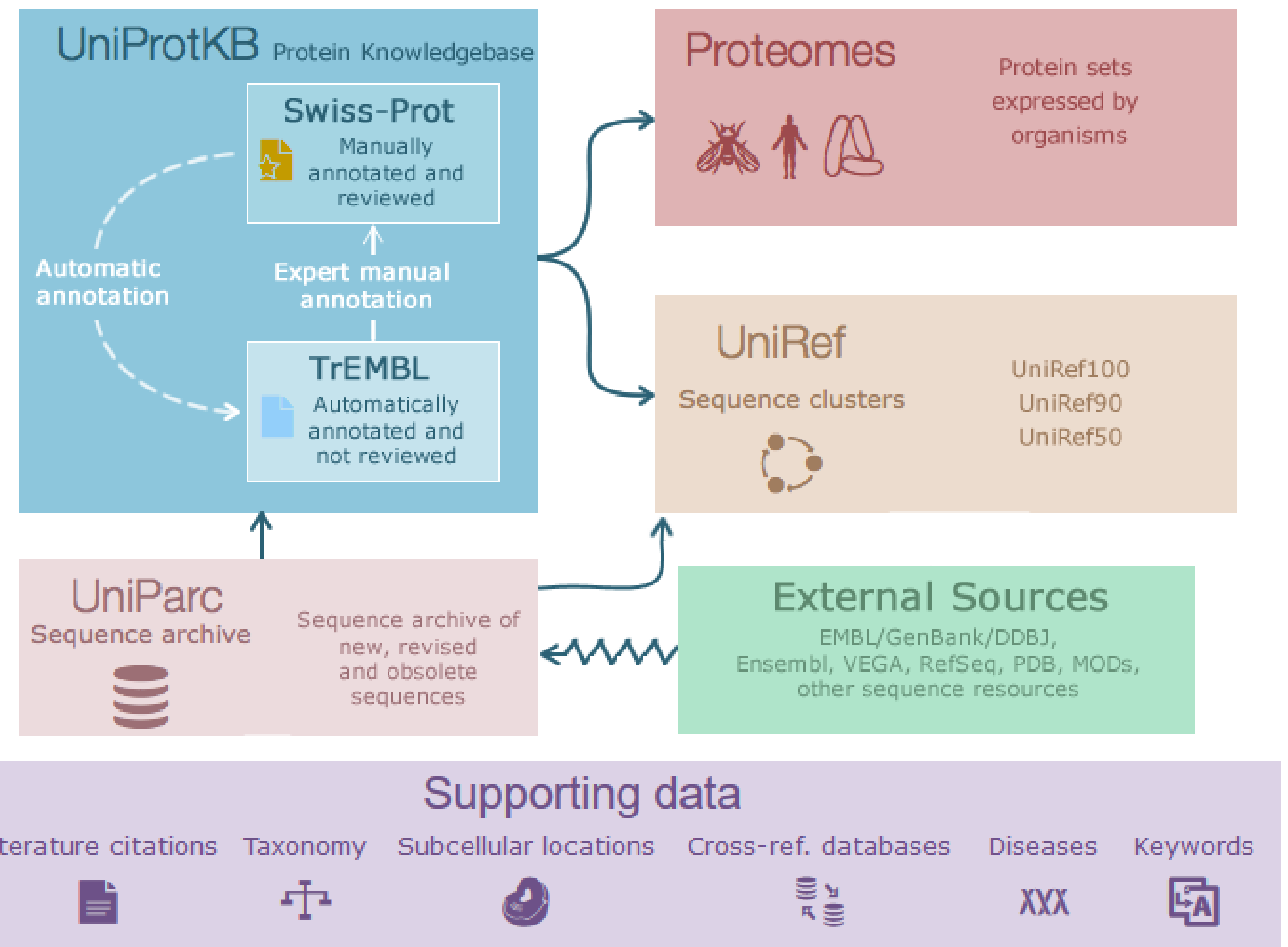
UniProt at a glance

The Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data. The UniProt databases are the UniProt Knowledgebase (UniProtKB), the UniProt Reference Clusters (UniRef), and the UniProt Archive (UniParc)

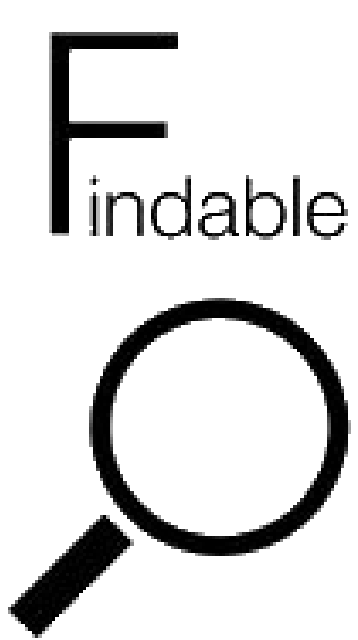
A FAIRer website



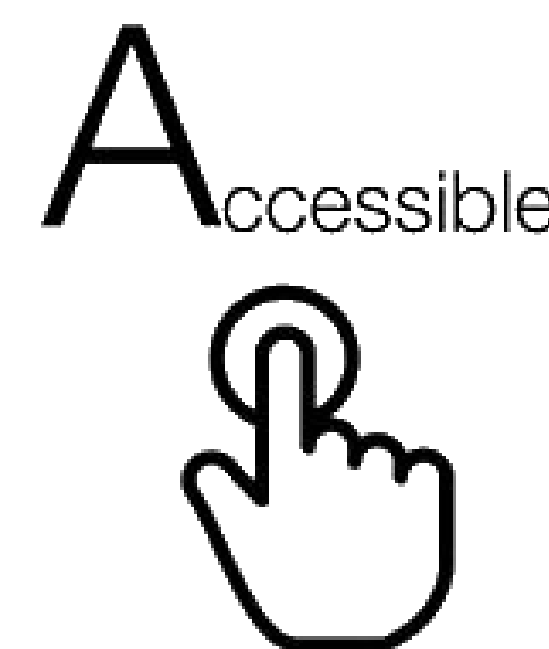
Website: moving from schema.org to Bioschemas



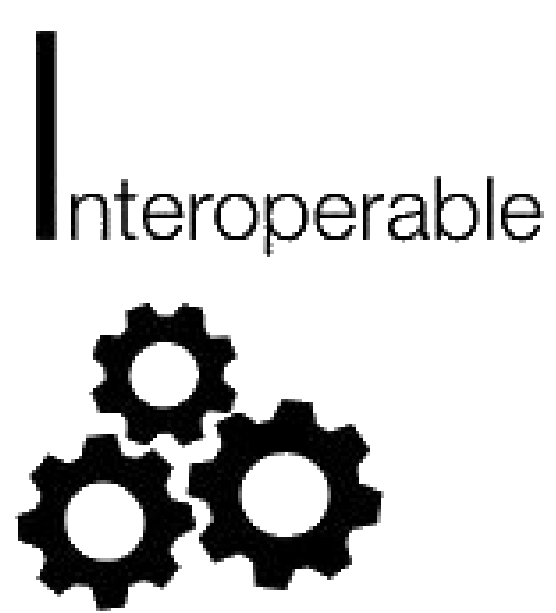
UniProt datasets, nearly 100% FAIR



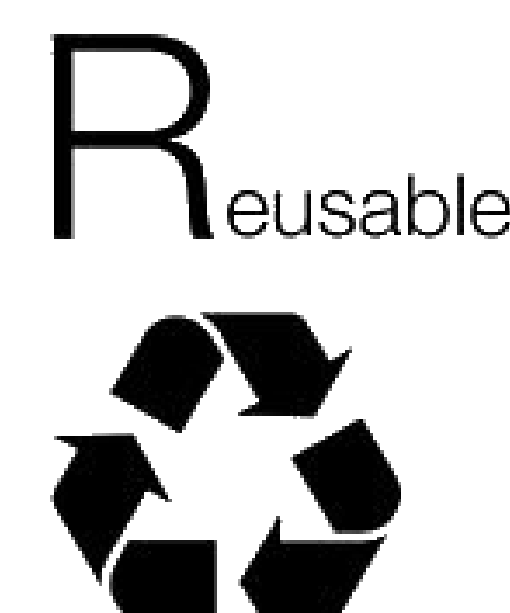
Metric	Assessment
FM-F1A Identifier uniqueness	Stable and unique identifiers → accessions
FM-F1B Identifier persistence	PURL identifiers
FM-F2 Machine readability of metadata	RDF format + VOID file FASTA files + headers XSD for XML
FM-F3 Resource identifier in metadata	PURL included in RDF Accession always provided
FM-F4 Indexed in a searchable resource	Indexed by Google, FAIRsharing, identifiers.org



Metric	Assessment
FM-A1.1 Access protocol	HTTPS for webpages & FTP for downloads
FM-A1.2 Access	Public datasets, not authorization required
FM-A2 Metadata longevity	Entry history always available Archive specialized dataset

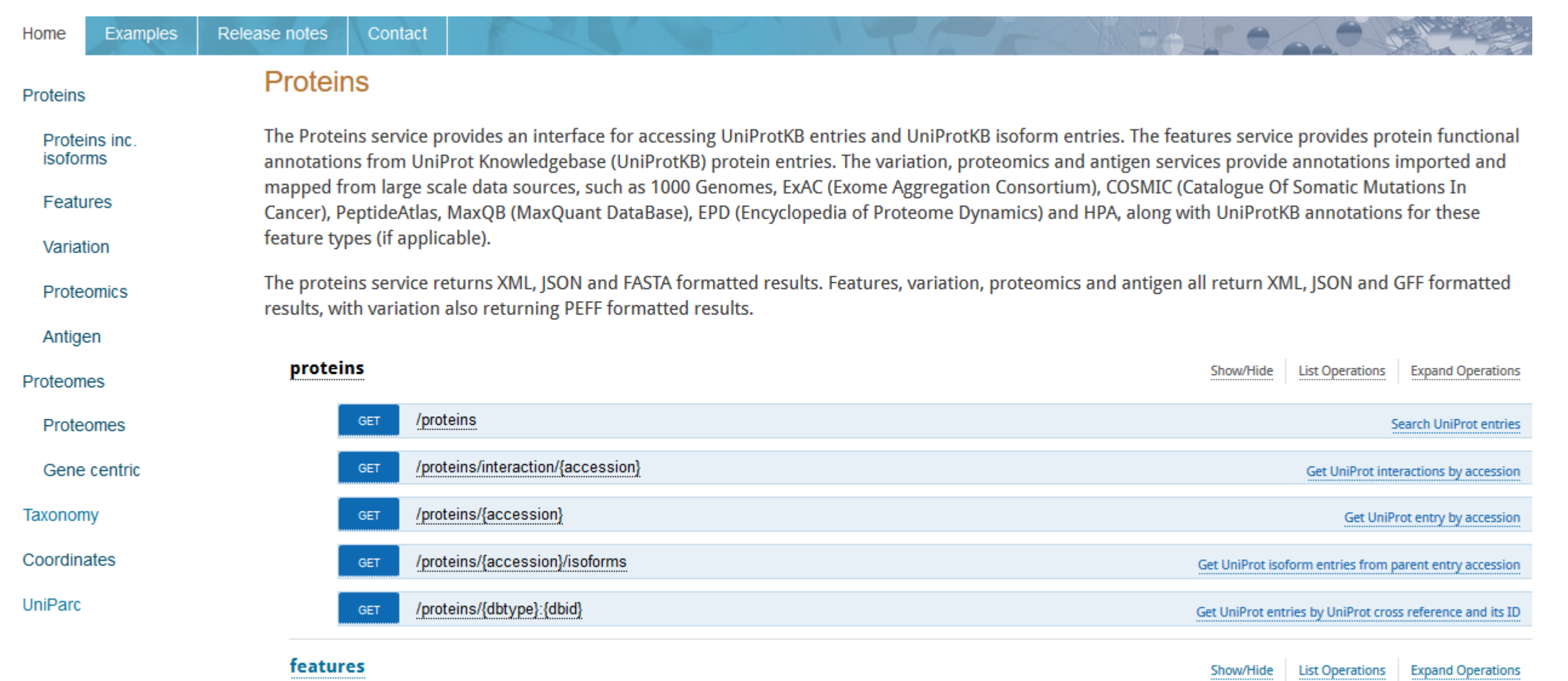
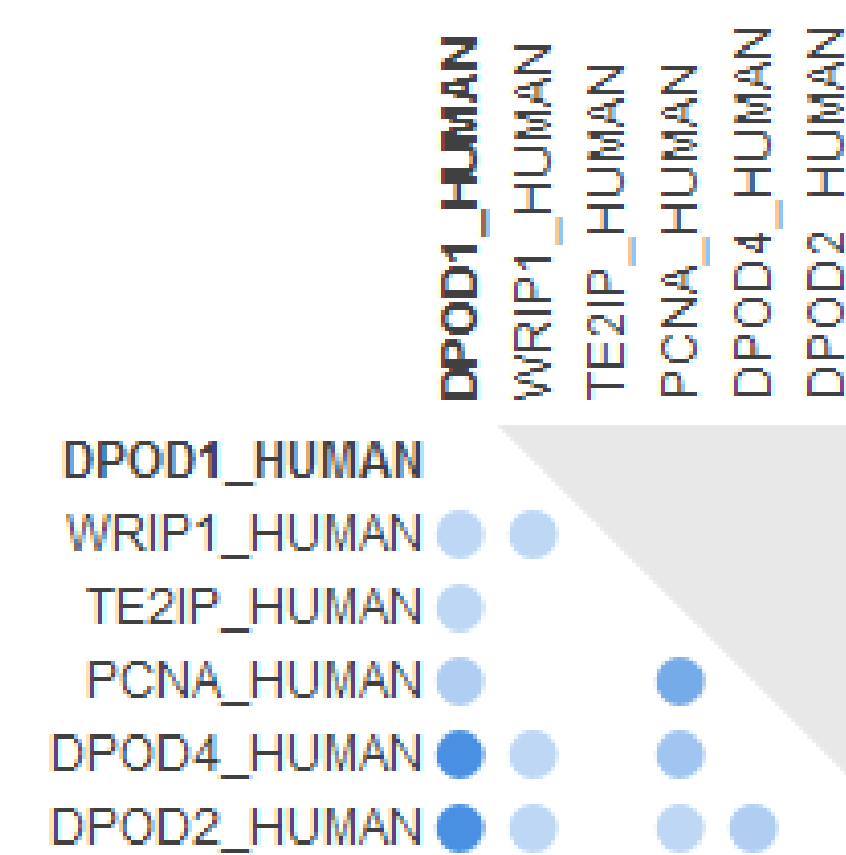


Metric	Assessment
FM-I1 Use a knowledge representation language	UniProt ontology for RDF
FM-I2 Use FAIR vocabularies	Uses well-known ontologies whenever possible
FM-I3 Use qualified references	"seeAlso" or "cross-reference" is commonly used



Metric	Assessment
FM-R1.1 Accessible usage license	Creative commons attribution (CC BY 4.0)
FM-R1.2 Detailed provenance	Provided on help pages Machine-readable via VOID for RDF On entries via Evidence Codes
FM-R1.3 Meets community standard	What certification body should be used?

UniProt tools looking for FAIRness (but no available metrics yet)



ProtVista and InteractionViewer → JavaScript projects

- F2. Basic metadata via package.json
- F4. Indexed at Google, GitHub and NPM
- I3. Modular architecture via web components with clear dependencies
- NPM packages
- GitHub ebi-webcomponents & ebi-uniprot
- R1.1 MIT or Apache
- R1.2 Basic provenance via package.json

Protein API → <https://www.ebi.ac.uk/proteins/api/doc/>

- F4. Indexed at Google and Bio.tools
- I1. JSON OpenAPI summary
- I3. Data linked to well-know ontologies whenever possible
- JSON, XML
- A1. REST service
- A1. HTTP
- Contact information on help pages

Funding
UniProt is supported by the National Institutes of Health (NIH), National Human Genome Research Institute (NHGRI) and National Institute of General Medical Sciences (NIGMS) grant U41HG007822. Additional support for the EMBL-EBI's involvement in UniProt comes from European Molecular Biology Laboratory (EMBL), the British Heart Foundation (BHF) (RG/13/5/30112), the Parkinson's Disease United Kingdom (PDUK) GO grant G-1307, and the NIH GO grant U41HG02273. UniProt activities at the SIB are additionally supported by the Swiss Federal Government through the State Secretariat for Education, Research and Innovation SERI. PIR's UniProt activities are also supported by the NIH grants R01GM080646, G08LM010720, and P20GM103446, and the National Science Foundation (NSF) grant DBI-1062520

<https://www.uniprot.org>

help@uniprot.org

<http://insideuniprot.blogspot.co.uk>

<http://www.ebi.ac.uk/training>