



Workbench Datasheet

Automated Data Cleansing and Standardization

@W0	rkbench Curate and ar	nnotate Search						*					:				
Sam	ple metac	lata							v 18, 2022, 15:16 ov 18, 2022, 15:16	\rightarrow	Annotati		~				
DATA	ANNOTATIONS	TOP UNANNOTATE	D SHARING REVI	SIONS								imers disease	_				
🐉 📽						Ann	otated cells: 61	98 of 22752 27%	Q		Alzheim INDIC/		2				
trow	Source Name	Characteristics[organism]	Characteristics[individual]	Characteristics[disease]		Annotation	NCI Annotation	Characteristics[disease staging]	Characteristic		5	U V T V					
1	↑ HB150318_13_003	T Homo saplens	个 AD7	Aizheimers disease	ICD10	↑ CM:G30.9	↑ NCI:C2866	↑ Braak stage III	frontal corte								
2	HB150318_13_005	Homo saplens	AD7	Alzheimers disease	ICD10	CM:G30.9	NCI:C2866	Braak stage III	frontal corte								
3	HB150318_13_007	Homo sapiens	AD7	Alzheimers disease	ICD10	@wa	orkbench cura	te and annotate	Search					•			:
4	HB150318_13_008	Homo saplens	AD7	Alzheimers disease	ICD10	Sam	nple me	etadata								Created: Nov 18, 20 Modified: Nov 18, 2	
5	HB150318_13_010	Homo saplens	AD7	Alzheimers disease	ICD10	DATA				ARING	REVI	SIONS			Annotated cells: 4302 of 2275		
6	HB150318_13_011	Homo saplens	AD7	Alzheimers disease	ICD10		Search	Ŧ	- Annotated p	oart		T Attribut	e Name	▼ Annotation source	Annotation method		ount
		R	8	14			Alzheimers d	isease	Alzheimer Disease	D00054		Characteristics	(disease)	INDICATION	MANUAL_VERIFIED	63	32
							frontal cortex		frontal cortex U000	1870		Characteristics	(organism part)	ANAT	MANUAL_VERIFIED 2	63	12
							MNase		Micrococcal Nuclea EFO0004177	ise dige	stion	Comment[LIBF	ARY_SELECTION]	EFO	MANUAL_VERIFIED	63	12
							Homo sapier	15	Homo sapiens TAX	DN9606		Characteristics	(organism)	EFO	MANUAL_VERIFIED	63	12
							female		Female PKPD10097	7		Characteristics	(sex)	PKPD	MANUAL_VERIFIED 2	63	32
Figure 1. Simple user interface and smart 🛛 🗛 🗛 🗛								Characteristics	(individual)	PROTEIN	MANUAL_VERIFIED	12	.19				
spreadsheet-like functionality for loading,									Factor Value(in	dividual]	PROTEIN	MANUAL_VERIFIED 🛃	12	:9			
emantically-annotating and cleaning						AD8 353128			Factor Value(in	dividual]	PROTEIN	MANUAL_VERIFIED	12	:8			
	ılar data		-	0			AD8		AD8 353128			Characteristics	(individual)	PROTEIN	MANUAL_VERIFIED	12	.8

- INTUITIVE Interactive sheet for fast and simple curation of data with terminology standards
- AUTOMATED Reproducible annotations using vocabs or public/private ontologies
- REPEATABLE Store annotations and share rules to reduce the time to curate new data.

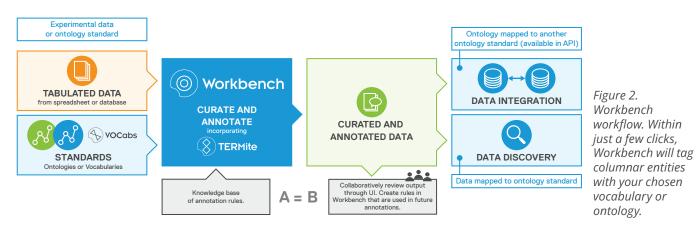
Take the effort out of tabular data curation

Recording data can pose organisations with numerous challenges. Individuals and groups may use idiomatic or historic nomenclature that limits data re-usability, and information can be siloed, restricting transparency and collaboration. Added to this, the process of integrating data from multiple sources can be time-consuming and prone to errors. SciBite has created Workbench, a simple visual interface for curating term lists, custom dictionaries, and semistructured datasets to your terminology standard of choice. Powered by SciBite's TERMite and VOCab technologies, SciBite Workbench serves to support organisations that are adopting a FAIR (Findable, Accessible, Interoperable, Reusable) approach to data management. A critical component to making data more FAIR is to enable data interoperability through aligning data to shared terminology and ontological standards.

Cleaning datasets by aligning them to ontologies can be an arduous task, often requiring specific expertise in both the subject domain and the available standards. Scientific data curators perform fundamental work within an organisation, enabling much of the downstream data integration and analysis. Workbench aims to support these scientists by streamlining the curation process through a simple and intuitive user interface. Workbench gives you the ability to reuse and repeat previously seen curation, saving time, and allowing teams to get more done (see figure 2).







Accurately curate complex data

Annotate data from selected columns with your chosen vocabulary or ontology, including SciBite's VOCabs - a library of manually curated vocabularies enriched with >20 million synonyms. You can also upload your own custom ontologies and use these for annotating your data.

Dealing with 'messy' data

Workbench includes SciBite's award winning Named Entity Recognition system, TERMite, that can be configured and fine tuned to support fuzzy-matching and handle variations in spelling and typographic errors. For data riddled with internal codes or proprietary terms, you can use Workbench to map them to your chosen ontology terms or vocabulary with 'annotation rules' eliminating the errorprone and monotonous process of editing. Workbench promotes replicability by storing these annotation rules, which can be re-run during subsequent data annotation tasks (see Figure 3).

Example 'annotation rules'						
Header context label	Cell value	Annotation				
gender/sex	1 exact	Female (PATO:0000383)				
gender/sex	F exact	Female (PATO:0000383)				
Human Cell Line	293HEK starts_with	HEK293 (CLO:0001230)				
species	Human ^{exact}	Homo Sapiens (NCBITaxon:9606)				
	The data cell value you wish to annotate	The terms Workbench will annotate your data to				

annotate your data to

Figure 3. Example of the types of annotation rules that can be replayed over your dataset.

Build workflows with the API

Workbench comes with a powerful REST API that provides programmatic access to the same core functions from the user interface. The API can be used to integrate Workbench functionality into your own custom data curation workflows.

Easy data sharing and export

Workbench has built-in sharing functionalities for collaborative curation projects. As a data owner, you can create a group where you can invite other colleagues to view or edit your annotations. Annotated data can be exported in Excel for use with many 3rd party tools.

Workbench Curate and annotate Se	sarch	T	1
Example E-MTAB-3732	2		Created: Nov 16, 2022, 14:55 Modified: Nov 16, 2022, 14:5
DATA ANNOTATIONS TOP UNANNO Owner: ADMIN.WB Group name: ADMIN.WB CHANGE PERMISSION ADD NEW USER REP	AVALUATE SHARING REVISIONS		Q
Username	Group role		
admin.wb	ADMIN		
ali	RO		
ali Dv	RO		

Figure 4. Add colleagues and collaborate on annotation tasks.

About SciBite

SciBite is an award-winning semantic software company offering an ontology-led approach to transforming unstructured content into machine-readable clean data. Supporting the top 20 pharma with use cases across life sciences, SciBite empowers customers with a suite of fast, flexible, deployable API technologies, making it a critical component in scientific data-led strategies. Contact us to find out how we can help you get more from your data.

Head Office: SciBite Limited **BioData Innovation Centre** Wellcome Genome Campus Hinxton, Cambridge CB10 1DR United Kingdom

 www.scibite.com contact@scibite.com in LinkedIn: SciBite Twitter: @SciBite +44 (0)1223 786 129