



Data format for Project Tycho 2.0 pre-compiled dataset CSV files

Version: 1.1

Date created: 2022-06-09

Project Tycho 2.0 pre-compiled datasets include counts of cases for conditions (i.e., disorders) in populations reported by health agencies. These counts can be specified for different geographical regions, for different time intervals, for various demographic groups and/or subpopulations, and can represent fatalities (deaths) or not. Each datafile includes these, and other, attributes of counts, as indicated below.

Missing data: Counts are often not reported for all time intervals and locations. Project Tycho 2.0 pre-compiled datasets only include time intervals and location with counts available and users will have to insert time intervals for which no counts were reported to complete time series of counts. Project Tycho 2.0 uses the value "NA" for missing information about count attributes.

Project Tycho 2.0 variables for pre-compiled datasets using data format v1.1

Attribute (* required)	Description	Value allowed	URL of vocabulary for attribute values
ConditionName*	Name of reported condition as listed in SNOMED-CT	Any SNOMED-CT disorder name	doi:10.25504/fairsharing.d88s6e
ConditionSNOMED*	SNOMED-CT code for reported condition	Any SNOMED-CT disorder code	doi:10.25504/fairsharing.d88s6e
PathogenName	Organism name for pathogen causing reported condition	Any organism name in the NCBI Taxonomy, or NA	doi:10.25504/fairsharing.fj07xj
PathogenTaxonID	NCBI Taxonomy identifier for pathogen causing reported condition	Any NCBI Taxonomy ID, or NA	doi:10.25504/fairsharing.fj07xj
Outcome	Condition outcome for which the count was reported	Any outcome listed by the NCI Thesaurus, including "Dead", "Complete recovery", or "Unspecified"	doi:10.25504/fairsharing.4cvwxa
MedicalProcedure	Medical or public health procedure for which the count was reported	"Hospital admission", "Admission to intensive care unit", "Unspecified"	doi:10.25504/fairsharing.d88s6e

Attribute (* required)	Description	Value allowed	URL of vocabulary for attribute values
CountryName*	ISO 3166 English Short Name of country for which a count was reported	Any ISO 3166 country English Short Name (all capitals), as of 2022-04-12	https://www.iso.org/standard/72482.html
CountryISO	ISO 3166 2-letter code for country	Any 2-letter ISO 3166 country code, as of 2022-04-12	https://www.iso.org/standard/63545.html
Admin1Name	ISO 3166 Name of first administrative subdivision for which a count was reported	Any ISO 3166 fist administrative subdivision name (all capitals), as of 2022-04-12, or NA. Boolean operators <AND>, <OR>, and <NOT> are permitted for combinations	https://www.iso.org/standard/63545.html
Admin1ISO	ISO 3166-2 code for first administrative subdivision	Any ISO 3166-2 code for first administrative subdivision, as of 2022-04-12, or NA	https://www.iso.org/standard/63545.html
Admin2Name	Placename from Geonames of second-order administrative division for which a count was reported	Any placename from Geonames of second-order administrative division (all capitals), or NA. Boolean operators <AND>, <OR>, and <NOT> are permitted for combinations	http://www.geonames.org/
CityName	Name of Populated place for which a count was reported, from Geonames	Any name from Geonames for (part of) a populated place, seat, capital, or similar feature class (all capitals), or NA. Boolean operators <AND>, <OR>, and <NOT> are permitted for combinations.	http://www.geonames.org/
SpecialRegionName	Name of a geographical region that is not part of the administrative subdivision hierarchy specified in Geonames, for which a count was reported	Any name of a geographical region. Boolean operators <AND>, <OR>, and <NOT> are permitted for combinations	No external standard

Attribute (* required)	Description	Value allowed	URL of vocabulary for attribute values
SpecialRegionURL	URL with the definition of the special region for which a count was reported	Any valid URL	No external standard
PeriodStartDate*	Start date of time interval for which a count was reported, in ISO 8601 format	Any date in ISO 8601 format	https://www.iso.org/iso-8601-date-and-time-format.html
PeriodEndDate*	End date of time interval for which a count was reported, in ISO 8601 format	Any date in ISO 8601 format. Must be no earlier than the PeriodStartDate	https://www.iso.org/iso-8601-date-and-time-format.html
PartOfCumulativeCountSeries*	Count is part of a series of cumulative counts (instead of being part of a series of fixed time interval counts)	0,1	No external standard
AgeRange*	Age range for which a count was reported (in years)	Any age range between 0 and 130 with age values separated by "-" as in: "0-130"	No external standard
Gender*	Gender identity (subjective sense of gender) for which a count was reported	Any gender listed by the NCI Thesaurus , or "All", or "Unspecified". Boolean operators <AND>, <OR>, and <NOT> are permitted for combinations	doi:10.25504/fairsharing.4cvwxa
BiologicalSex*	Biological sex (most often assigned at birth based on phenotype) for which a count was reported	"Female", "Male", "All", or "Unspecified". Boolean operators <AND>, <OR>, and <NOT> are permitted for combinations	doi:10.25504/fairsharing.ezwdhz
EthnicGroup*	Ethnic group for which the count was reported	Any group listed in the NCI Thesaurus OBO Edition or "Unspecified". Boolean operators <AND>, <OR>, and <NOT> are permitted for combinations	doi:10.25504/fairsharing.4cvwxa
Race*	Race for which the count was reported	Any race listed in the NCI Thesaurus OBO Edition or	doi:10.25504/fairsharing.4cvwxa

Attribute (* required)	Description	Value allowed	URL of vocabulary for attribute values
		"Unspecified". Boolean operators <AND>, <OR>, and <NOT> are permitted for combinations	
Subpopulation*	A subgroup of the population for which a count was reported	"Civilian", "Military", "Military - Veterans", "Cruise ship - Diamond Princess", "Cruise ship - Grand Princess", "Prison population - Federal", "Prison population - State", Any other subpopulation, or "Unspecified". Boolean operators <AND>, <OR>, and <NOT> are permitted for combinations	No external standard
PlaceOfAquisition	Place where reported condition was acquired: "Domestic" (i.e., in the country for which a case count was reported) or "Abroad" (i.e., not in the country for which the case count was reported)	"Domestic", "Abroad", or NA	No external standard
DiagnosisCertainty	Qualifier for certainty of diagnosis for a count condition	"Definite", "Equivocal", "Possible diagnosis", "Probable diagnosis", or NA. Boolean operators <AND>, <OR>, and <NOT> are permitted for combinations	doi:10.25504/fairsharing.d88s6e
SourceName*	Name of the source (system, database, institution) from which counts were obtained by the Project Tycho team	E.g., "Nationally Notifiable Disease Surveillance System", "World Health Organization DengueNet Database"	No external standard
CountValue*	The count value	Any whole number	No external standard

Example of part of a dataset:

ConditionName	ConditionSNOMED	PathogenName	PathogenTaxonID	Outcome	MedicalProcedure	CountryName	CountryISO	Admin1Name	Admin1ISO	Admin2Name	CityName	SpecialRegionName	SpecialRegionURL	PeriodStartDate	PeriodEndDate	PartOfCumulativeCountSeries	AgeRange	Gender	BiologicalSex	EthnicGroup	Race	Subpopulation	PlaceOfAcquisition	DiagnosisCertainty	SourceName	CountValue
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Dead	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	AUTAUGA COUNTY	NA	NA	NA	2020-01-22	2020-03-31	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	0
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Dead	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	AUTAUGA COUNTY	NA	NA	NA	2020-01-22	2020-04-01	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	7
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Unspecified	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	AUTAUGA COUNTY	NA	NA	NA	2020-01-22	2020-03-31	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	0
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Dead	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	CLARKE COUNTY	NA	NA	NA	2020-01-22	2020-03-31	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	0
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Dead	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	CLARKE COUNTY	NA	NA	NA	2020-01-22	2020-04-01	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	10
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Unspecified	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	CLARKE COUNTY	NA	NA	NA	2020-01-22	2020-03-31	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	0
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Unspecified	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	CLARKE COUNTY	NA	NA	NA	2020-01-22	2020-04-01	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	2
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Dead	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	DALLAS COUNTY	NA	NA	NA	2020-01-22	2020-03-31	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	0
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Dead	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	DALLAS COUNTY	NA	NA	NA	2020-01-22	2020-04-01	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	0
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Unspecified	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	DALLAS COUNTY	NA	NA	NA	2020-01-22	2020-03-31	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	2
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Unspecified	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	DALLAS COUNTY	NA	NA	NA	2020-01-22	2020-04-01	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	3
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Dead	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	JEFFERSON COUNTY	NA	NA	NA	2020-01-22	2020-03-31	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	0
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Dead	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	JEFFERSON COUNTY	NA	NA	NA	2020-01-22	2020-04-01	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	11
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Unspecified	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	JEFFERSON COUNTY	NA	NA	NA	2020-01-22	2020-03-31	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	277
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Unspecified	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	JEFFERSON COUNTY	NA	NA	NA	2020-01-22	2020-04-01	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	305
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Dead	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	LEE COUNTY	NA	NA	NA	2020-01-22	2020-03-31	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	2
COVID-19	840539006	Severe acute respiratory syndrome coronavirus 2	2697049	Dead	Unspecified	UNITED STATES OF AMERICA	US	ALABAMA	US-AL	LEE COUNTY	NA	NA	NA	2020-01-22	2020-04-01	1	0-130	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	NA	NA	Alabama Department of Health Website Dashboard	75

In case of questions, please visit www.tycho.pitt.edu or email tycho@phdl.pitt.edu