HDRUK
Health Data Research UK

Descriptive Metadata Specification V2

12.8.2020

Purpose

The following document has been developed to refine and iterate the metadata needed for the Innovation Gateway.

The specification is initially focused on Descriptive metadata (see proposed metadata levels in principles below).

Document Control Information							
Version	Status	Author(s)	Date Created	Reviewer(s)	Date Reviewed	Approved	Notes
V2.0.0	Release Candidate	A.Milward, D.Milward	05/08/2020	N.Sebire	11.8.2020	11.8.2020	
				B.Gordon	11.8.2020	11.8.2020	
				S.Varma	11.8.2020	11.8.2020	
Minimum Retention Period			Permanent				
Disposal Required			Archive when new version created				
Classification			PUBLIC				
Next Review due:			12/011/2020				

Contents

PURPOSE 2
CONTENTS 3
INTRODUCTION 5
Layout: 6
KEY TERMS: 7
DESCRIPTIVE METADATA 9
Identifier 9
Version 9
Issued 9
Modified 10
REVISIONS 10
Version 10
Url 10
SUMMARY 11
Title 11
Abstract 11
Contact Point 12
Keywords 13
DoiName 13
PUBLISHER 14
Identifier 14
Name 15
Logo 15
Description 16
Contact Point 16
Member Of 17
Delivery Lead Time 18
Access Rights 18
Access Service 19
Access Request Cost 20
Data Use Limitation 20
Data Use Requirements 21
Documentation 22
Description 22
Associated Media 23
Is Part Of 23
Coverage 24
Spatial 24
Typical Age Range 25
Physical Sample Availability 25
Followup 26
Pathway 26
Provenance 27
ORIGIN 27
Purpose 27
Source 27
Collection Situation 28
TEMPORAL 29
Accrual Periodicity 29
Distribution Release Date 30
Start Date 30
End Date 31
Time Lag 32
Accessibility 32
Usage 32
Data Use Limitation 32
Data Use Requirements 33
Resource Creator 34
Investigations 34
Is Referenced By 34

Access *35*
Access Rights 35
Access Service 36
Access Request Cost 37
Delivery Lead Time 37
Jurisdiction 38
Data Controller 39
Data Processor 39
Formats & Standards *40*
Vocabulary Encoding Scheme 40
Conforms To 42
Language 43
Format 44
Enrichment & Linkage *45*
Qualified Relation 45
Derivation 46
Tools 46
Observations *47*
Observation *47*
Observed Node 48
Measured Value 48
Disambiguating Description 49
Observation Date 49
Measured Property 49
APPENDIX 1: CHANGE LOG **51**

Introduction

As part of the continued development of the Innovation Gateway Minimum Viable Product (MVP), Health Data Research UK (HDR UK) will continue to collect and improve the information about datasets that are available for research and innovation across their alliance member and their health data research hubs. This information must be human and machine-readable, allowing researchers to understand the data that is available, as well as allowing the Gateway to index the information and return relevant search results.

V2.0.0 builds on feedback from the initial specification (V1.1.7) and its implementation during the Innovation Gateway MVP phase. In addition, we've engaged academic, life science, industry and public sector data users and providers, as part of wider consultation that seeks to address which metadata attributes will allow users to evaluate the utility ("fit for purpose x") of a dataset.

The following principles have governed the work and decisions made:

- Metadata should be based on a user need and a user story
- Metadata should be based on standard terminologies whenever possible including:
    - DCAT
    - Schema.org
    - Dublin Core
    - DDI
- If an identified need does not have an equivalent entry from the standards above, another relevant standard should be selected
- Only if no standard can be found should we create a new metadata entry
- Naming convention: Metadata "titles" will be lowercase, underscore separated (note that in the human readable specification we are using Capitalised for ease of reading)
- Where standard attributes have been used, we will use the name from that standard. However, we may 'Label' the attributes and provide guidance that makes the attribute more user friendly as many of our users will not be familiar with the more technical terminology.
- Prioritisation of Metadata should be based on Impact vs Effort (effort needs to consider potential for automation in the longer-term vs manual in nearer term).

As part of a wider scope of work we are seeking to work with the British Standards Institute, UK Data Standards Authority and the International Standards Organisation to gain wider adoption and increase metadata interoperability across organisations using this Healthcare Informatics Profile. A more generic iteration of this document may be created for these use cases.

This document only seeks to address Descriptive Metadata. A subsequent document will be published that seeks to address structural metadata within the wider context of Metadata Federation and Interoperable Data/Metadata catalogues.

Layout:

The document has been laid out in the following way:

Metadata element name

Primitive data type

Logo	Please provide a logo associated with the Gateway Organisation using a valid URI. The following formats will be accepted .jpg, .png or .svg. **Note**: the image asset can be hosted by the organisation or uploaded using the onboarding portal.	xs:string //Regular expression (regex) to validate URL: x==~ /https?:\/\/(www\.)?[-a-zA-Z0-9@:%._\+~#=]{1,256}\.[a-zA-Z0-9()]{1,6}\b([-a-zA-Z0-9()@:%_\+.~#?&//=]*)/

Regular Expression (v3). Java.util.regex api will be used to validate entries. See for more info.

Completion guidance for the Metadata

Some metadata elements may have enumerated values rather than regular expressions to constrain the values provided:

Membership	Please indicate if the organisation is an Alliance Member or a Hub.	xs:string HUB: https://www.hdruk.ac.uk/infrastructure/the-hubs/ ALLIANCE: https://ukhealthdata.org/

Constrained list of values

Key terms:

Digital Innovation Hub: Initially funded through the UK Research and Innovation Industrial Strategy Challenge Fund (ISCF) investment to kick-start the approach, these Hubs are formal collaborations between NHS, academic organisations, patients, charities and industry. Each Hub will be set up to demonstrate and drive the utility of expert health data science, based on an area of expertise. But this is not a research project – Hubs will be expected to provide data, curate the data and offer services for research and innovation. Please see https://www.hdruk.ac.uk/infrastructure/the-hubs/ for more information.

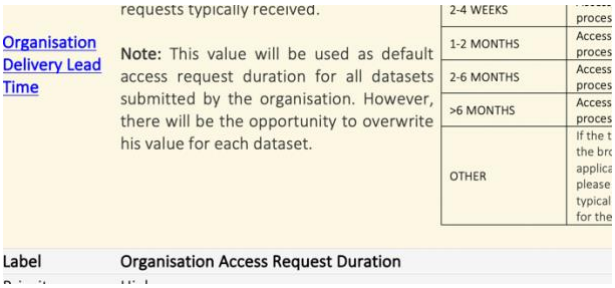
Alliance: The UK Health Data Research Alliance is an independent alliance of leading healthcare and research organisations united to establish best practice for the ethical use of UK health data for research at scale. Its members include organisations representing national organisations, NHS trusts, research institutes and charities. Please see https://ukhealthdata.org/members/ for more information

Innovation gateway MVP: Created with input from patients and researchers, this first phase of the Gateway provides a library of information about over 400 health datasets that are held and managed by the NHS, research charities, research institutes and universities. Please see https://healthdatagateway.org/ for more information.

HDRUK onboarding platform (HOP): Created to address the challenge of onboarding metadata from Alliance Members and Health Data Research Hubs the HOP provides utilities that allow organisations to submit metadata and for HDRUK to track submissions. Please see https://hdruk.metadata.works/ for more information.

Dataset: We are using the schema.org definition of a dataset as a "body of structured information describing some topic(s) of interest". Each data published is asked to reasonably evaluate what they consider a discrete dataset within the context of research use and discovery.

Label: In some cases, labels for attributes will be different from the underlying names used to store and interchange the information.



Where possible, the name of the attribute is consistent with the underlying standard i.e. schema.org , whilst the label is representative of the user experience when they input the information. Note that for a few attributes, we have reverted to the 'standard' name, but kept the label from previous versions.

User Story: Illustrates the purpose for collecting this metadata

As A [user]	Data custodian / researcher / coordinator
I Want To	Know the organisation responsible for submitting and managing the metadata through the HDR UK Gateway
So That	Identify who is responsible for the metadata

Example: Illustrative values that could be provided:

Example	https://www.hdruk.ac.uk/wp-content/themes/hdruk/assets/img/logo.svg

Priority: Indication of the priority of the metadata. Can be High, Medium or Low:

Priority	High

Completion Guidance: Metadata that has a min occurrence of >=1 or * is mandatory (* indicates many). If there is a min occurrence of 0 then the field is not mandatory. Some metadata may not be mandatory because it isn't applicable to an organisation, but completion is still be required, if it is relevant. Some metadata may not be mandatory, but completion is recommended.

Completion	Mandatory
Cardinality	Min Occurs: 1 Max Occurs: 1

Definition: Formal definition for the metadata, usually referencing a standard.

Definition	The name of the item. Canonical URL: http://schema.org/name Equivalent Property: dct:title

Notes for HOP: These are specific notes for the HDRUK Onboarding platform and Onboarding team. However, some of these notes are relevant to those submitting data so they have been included for reference.

Notes for HOP	RGCP	https://grid.ac/institutes/grid.451233.2
	Genomics England	https://grid.ac/institutes/grid.498322.6
	DATA-CAN	None
	Gut Reaction	None
	INSIGHT	None
	DiscoverNOW	None
	Cystic Fibrosis Trust	https://grid.ac/institutes/grid.453642.2
	NIHR BioResource	Many sites. None for the central function.
	NSS Scotland	https://www.grid.ac/institutes/grid.422655.2

Schema.org alternative: If available, a schema.org alternative has been identified. This will allow us to provide structured information for search engines. The HOP will publish the information in JSON-LD for 'LIVE' models.

Schema.org alternative	Schema.org Dataset/name

Survey metadata: Alliance members and Hubs were surveyed and asked to evaluate the attributes collected as part of V1 of the metadata specification. We have included the results of the survey for each element to inform the consultation process for V2. Note that some elements were not part of V1 so these will not have survey results.

Descriptive Metadata

Describes a resource for purposes of discovery and identification.

Identifier	Resolvable, Persistent Dataset Identifier This is the persistent identifier that should resolve to the latest metadata revision.	urlOrUUIDv4 // value is uri or uuid x ==~ /^((([A-Za-z]{3,9}:(?:\/\/)?)(?:[-;:&=\+\$,\w]+@)?[A-Za-z0-9.-]+(:[0-9]+)?|(?:www.|[-;:&=\+\$,\w]+@)[A-Za-z0-9.-]+)((?:\/[\+~%\/.\w-_]*)?\??(?:[-\+=&;%@.\w_]*)#?(?:[\w]*))?)|([a-fA-F0-9]{8}-[a-fA-F0-9]{4}-[a-fA-F0-9]{4}-[a-fA-F0-9]{4}-[a-fA-F0-9]{12})$/
Label	Resolvable Identifier	
Priority	High	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Note:	The identifier might be used as part of the URI of the item, but still having it represented explicitly is useful.	
Definition	A unique identifier of the item. RDF Property: dct:identifier Range: rdfs:Literal Source: dcat:resource_identifier Vocabulary: Data Catalog Vocabulary (DCAT)	
As A [user]	Researcher	
I Want To	Know the local identifier for the dataset	
So That	I can identify or refer to it	
Survey: Metadata Needed? (5 = Definitely Needed)	4.38	
Survey: Burden (5 = Easy)	4.43	

Version	Dataset metadata version	semver / x.x.x x ==~ /^([0-9]+)\.([0-9]+)\.([0-9]+)$/
Example	1.1.0	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Priority	High	
Title	Metadata Version	
Vocabulary	Dct:hasVersion	

Issued	Dataset Metadata Date Published Date.	xs:dateTime

Title	Date Published	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Vocabulary	Dcat:issued	
Priority	High	

Modified	Dataset Metadata Modification Date	xs:dateTime
Title	Date Last Modified	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Vocabulary	dcat:modified	
Priority	High	

Revisions

Revisions of Dataset metadata. This is a repeating section where all the revisions (versions) of the metadata can be captured and linked.

Version	Dataset Metadata version	semver / x.x.x x ==~ /^([0-9]+)\.([0-9]+)\.([0-9]+)$/
Cardinality	Min Occurs: 1 Max Occurs: 1	
Priority	High	
Vocabulary	dct:hasVersion	

Url	URL endpoint to obtain the version	url ///Regular expression (regex) to validate URL: x ==~ /((([A-Za-z]{3,9}:(?:\/\/)?)(?:[-;:&=\+\$,\w]+@)?[A-Za-z0-9.-]+(:[0-9]+)?|(?:www.|[-;:&=\+\$,\w]+@)[A-Za-z0-9.-]+)((?:\/[\+~%\/.\w-_]*)?\??(?:[-\+=&;%@.\w_]*)#?(?:[\w]*))?)/
Cardinality	Min Occurs: 1 Max Occurs: 1	
Vocabulary	dcat:distribution	

Summary

Summary metadata must be completed by Data Custodians onboarding metadata.

Title	Name of the dataset limited to 80 characters. It should provide a short description of the dataset and be unique across the gateway. If your title is not unique, please add a prefix with your organisation name or identifier to differentiate it from other datasets within the Gateway. Please avoid acronyms wherever possible. Good titles should summarise the content of the dataset and if relevant, the region the dataset covers. **Notes**: if the dataset is a "linked dataset", please indicate this using the prefix "Linked".	80characters //Regular expression (regex) to validate and constrain title between 2 and 80 characters x==~ /^.{2,80}$/
Examples	North West London COVID-19 Patient Level Situation Report	
Priority	High	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Definition	A name given to the item. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dct:title Range: rdfs:Literal Source: dcat:resource_title	
As A [user]	Researcher	
I Want To	Know the name of the dataset	
So That	I can identify and find it	
Schema.org alternative	Schema.org Dataset/name	
Survey: Metadata Needed? (5 = Definitely Needed)	4.75/5	
Survey: Burden (5 = Easy)	4.71/5	

Abstract	Provide a clear and brief descriptive signpost for researchers who are searching for data that may be relevant to their research. The abstract should allow the reader to determine the scope of the data collection and accurately summarise its content. The optimal length is one paragraph (limited to 255 characters) and effective abstracts should avoid long sentences and abbreviations where possible. **Note**: Researchers will view Titles and Abstracts when searching for datasets and choosing whether to explore their content further. Abstracts must be different from the full description for a dataset.	abstractText //Regular expression (regex) to validate and constrain title between 5 and 255 characters x==~ /^.{5,255}$/
Example	CPRD Aurum contains primary care data contributed by General Practitioner (GP) practices using EMIS Web® including patient registration information and all care events that GPs have chosen to record as part of their usual medical practice.	

Priority	High
Completion	Mandatory
Cardinality	Min Occurs: 1 Max Occurs: 1
Definition	A summary of the resource. Vocabulary: DCMI Metadata Terms RDF Type: rdf:Property Refines: dc:description Source: dct:abstract
As A [user]	Researcher
I Want To	read a plain English summary of the dataset
So That	I can identify if it is of interest
Schema.org alternative	Schema.org Dataset/abstract
Survey: Metadata Needed? (5 = Definitely Needed)	4.5/5
Survey: Burden (5 = Easy)	4/5

Contact Point	Please provide a valid email address that can be used to coordinate data access requests with the publisher. Organisations are expected to provide a dedicated email address associated with the data access request process. **Notes**: An employee's email address can only be provided on a temporary basis and if one is provided an explicit consent must be obtained for this purpose.	email //Regular expression (regex) to validate email addresses x ==~/[^\s]+@[^\s]+\.[^\s]+/
Default	Organisation contact point	
Examples	SAILDatabank@swansea.ac.uk	
Priority	High	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Definition	Relevant contact information for the catalogued resource. Use of vCard is recommended [VCARD-RDF] . Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dcat:contactPoint Range: vcard:Kind Source: dcat:resource_contact_point	
As A [user]	Researcher	
I Want To	Know the contact point	
So That	I can contact them about the dataset and request access	
Schema.org alternative	https://schema.org/email	
Notes for HOP	At registration organisations will have the opportunity to add a default contact point for their organisation. If users would like to change default values, they will be given the opportunity to do so by 'un-checking' a tick box next to the item.	

Survey: Metadata Needed? (5 = Definitely Needed)	4.88/5
Survey: Burden (5 = Easy)	4.14/5

Keywords	Please provide relevant and specific keywords that can improve the SEO of your dataset as a comma separated list. **Notes:** Onboarding portal will suggest keywords based on title, abstract and description. We are compiling a standardised list of keywords and synonyms across datasets to make filtering easier for users. These keywords will be split and stored as an array.	80characters x==~ /^.{2,80}$/
Default	Health Data, Research	
Examples	SAIL, Primary care, GP	
Priority	Medium	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: *	
Definition	A keyword or tag describing the resource. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dcat:keyword Range: rdfs:Literal Source: dcat:resource_keyword	
As A [user]	Researcher	
I Want To	Identify the keywords associated with the data	
So That	I can index and search data, as well as look up similar datasets	
Survey: Metadata Needed? (5 = Definitely Needed)	4/5	
Survey: Burden (5 = Easy)	3.57/5	

DoiName	All HDR UK registered datasets should either have a Digital Object Identifier (DOI) or be working towards obtaining one. If a DOI is available, please provide the DOI. You can either use the HDR UK Onboarding Portal (HOP) to mint a DOI for your dataset or your chosen alternative means. **Note:** This is not the DOI of the publication(s) associated with the dataset.	doi //Regular expression (regex) to validate DOI from https://www.crossref.org/blog/dois-and-matching-regular-expressions/ x==~ /^10.\d{4,9}/[-._;()/:A-Z0-9]+$/i
Priority	High	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Definition	Digital Object Identifier. Provides an actionable, interoperable, persistent link Actionable – through use of identifier syntax and network resolution mechanism (Handle System®) Persistent – through combination of supporting improved handle infrastructure (registry database, proxy support, etc) and social infrastructure (obligations by Registration Agencies)	

	Interoperable – through use of a data model providing semantic interoperability and grouping mechanisms. Vocabulary: DOI Data Dictionary Source: DOI
As A [user]	Data custodian / researcher
I Want To	Provide a persistent interoperable identifier for a dataset
So That	I can uniquely identify it, globally, and provide a HTTP URI to navigate to metadata that describes it.
Notes for HOP	HOP will mint DOIs
Survey: Metadata Needed? (5 = Definitely Needed)	3.63 / 5
Survey: Burden (5 = Easy)	4.43/5

Publisher

Completion Notes: This is the organisation responsible for running or supporting the data access request process, as well as publishing and maintaining the metadata.

In most this will be the same as the HDR UK Organisation (Hub or Alliance Member).

However, in some cases this will be different i.e. Tissue Directory are an HDR UK Gateway organisation but coordinate activities across a number of data publishers i.e. Cambridge Blood and Stem Cell Biobank.

Information about each organisation onboarding and curating metadata for a given dataset must be provided, at registration, including:

Identifier	Please provide a Grid.ac identifier (see https://www.grid.ac/institutes) for your organisation If your organisation does not have a Grid.ac identifier please use the "suggest and institute" function here: https://www.grid.ac/institutes#	url //Regular expression (regex) to validate URL: x ==~ /((([A-Za-z]{3,9}:(?:\/\/)?)(?:[-;:&=\+\$,\w]+@)?[A-Za-z0-9.-]+(:[0-9]+)?|(?:www.|[-;:&=\+\$,\w]+@)[A-Za-z0-9.-]+)((?:\/[\+~%\/.\w-_]*)?\??(?:[-\+=&;%@.\w_]*)#?(?:[\w]*))?)/
Priority	Medium	
Completion	Required	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Notes for HOP	RGCP: https://grid.ac/institutes/grid.451233.2 Genomics England: https://grid.ac/institutes/grid.498322.6 DATA-CAN: None Gut Reaction: None	

	INSIGHT	None
	DiscoverNOW	None
	Cystic Fibrosis Trust	https://grid.ac/institutes/grid.453642.2
	NIHR BioResource	Many sites. None for the central function.
	NSS Scotland	https://www.grid.ac/institutes/grid.422655.2
	Tissue Directory	None
	SAIL	None
	BREATHE	None
	NHS Digital	https://www.grid.ac/institutes/grid.498467.0
	Digitrial	None
	HQIP	https://www.grid.ac/institutes/grid.484216.d
	CPRD	https://www.grid.ac/institutes/grid.477301.6
	Barts Health NHS Trust	https://www.grid.ac/institutes/grid.139534.9
	University Hospitals Leicester NHS Trust	https://www.grid.ac/institutes/grid.269014.8
	Oxford	https://www.grid.ac/institutes/grid.410556.3
	PIONEER	None
Definition:	Global Research Identifier Database URI	
As A [user]	Data custodian / researcher / coordinator	
I Want To	Access standardised metadata about the organisation	
So That	I can understand the context of the organisation	

Name	Please provide the name of the Hub or Alliance member.	80characters //Regular expression to validate min, max character limit x==~ /^.{2,80}$/
As A [user]	Data custodian / researcher / coordinator	
I Want To	Know the organisation responsible for submitting and managing the metadata through the HDR UK Gateway	
So That	I can identify who is responsible for the metadata	
Priority	High	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Definition	The name of the item. Canonical URL: http://schema.org/name Equivalent Property: dct:title	

Logo	Please provide a logo associated with the Gateway Organisation using a valid URL. The following formats will be accepted .jpg, .png or .svg. **Note**: the image asset can be hosted by the organisation or uploaded using the onboarding portal.	url //Regular expression (regex) to validate URL: x==~ /https?:\/\/(www\.)?[-a-zA-Z0-9@:%._\+~#=]{1,256}\.[a-zA-Z0-

		9()]{1,6}\b([-a-zA-Z0-9()@:%_\+.~#?&//=]*)/
Example	https://www.hdruk.ac.uk/wp-content/themes/hdruk/assets/img/logo.svg	
Priority	High	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Definition	An associated logo. Canonical URL: http://schema.org/logo	
As A [user]	Data custodian / researcher / coordinator	
I Want To	Know the organisation responsible for submitting and managing the metadata through the HDR UK Gateway	
So That	I can identify who is responsible for the metadata	

Description	Please provide a URL that describes the organisation. **Note**: Hubs can include a link to their HDR UK listing, Alliance members can include a link to their ukhealthdata.org page	url //Regular expression (regex) to validate URL: x ==~ /((([A-Za-z]{3,9}:(?:\/\/)?)(?:[-;:&=\+\$,\w]+@)?[A-Za-z0-9.-]+(:[0-9]+)?|(?:www.|[-;:&=\+\$,\w]+@)[A-Za-z0-9.-]+)((?:\/[\+~%\/.\w-_]*)?\??(?:[-\+=&;%@.\w_]*)#?(?:[\w]*))?)/
Examples	https://www.hdruk.ac.uk/infrastructure/the-hubs/breathe/ https://ukhealthdata.org/members/barts-health-nhs-trust/`	
Priority	High	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Definition	A description of the organisation Canonical URL: http://schema.org/description	
As A [user]	Data custodian / researcher / coordinator	
I Want To	Know more about the organisation responsible for submitting and managing the metadata through the HDR UK Gateway	
So That	I can understand the context of the data and the submission	

Contact Point	Please provide a valid email address that can be used to coordinate data access requests with the publisher. Organisations are expected to provide a dedicated email address associated with the data access request process for their organisation. This will then be used as a default contact point for all datasets submitted by the organisation. However, there will be the opportunity to overwrite his value for each dataset.	email //Regular expression (regex) to validate email addresses x ==~/[^\s]+@[^\s]+\.[^\s]+/

	Notes: An employee's email address can only be provided on a temporary basis and if one is provided an explicit consent must be obtained for this purpose.
Examples	SAILDatabank@swansea.ac.uk
Priority	High
Completion	Recommended
Cardinality	Min Occurs: 1 Max Occurs: 1
Definition	Relevant contact information for the catalogued resource. Use of vCard is recommended [VCARD-RDF] . Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dcat:contactPoint Range: vcard:Kind Source: dcat:resource_contact_point
As A [user]	Researcher
I Want To	Know the contact point
So That	I can contact them about the dataset and request access
Schema.org alternative	https://schema.org/email
Notes for HOP	At registration organisations will have the opportunity to add a default contact point for their organisation. If users would like to change default values, they will be given the opportunity to do so by 'un-checking' a tick box next to the item. If the defaults are updated, then all datasets that have not been 'un-checked' will be overwritten with the new value.
Survey: Metadata Needed? (5 = Definitely Needed)	4.88/5
Survey: Burden (5 = Easy)	4.14/5

Member Of	Please indicate if the organisation is an Alliance Member or a Hub.	membership HUB: https://www.hdruk.ac.uk/infrastructure/the-hubs/ ALLIANCE: https://ukhealthdata.org/ OTHER: Other organisation affiliated with HDR UK
Title	Organisation Membership	
Priority	High	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Definition	An indicator of the organisations HDR UK membership	
As A [user]	Data custodian / researcher / coordinator	
I Want To	Know if the organisation is a Hub or an Alliance Member	
So That	I can understand the context of the data and the submission	

Delivery Lead Time	Please provide an indication of the typical processing times based on the types of requests typically received. **Note:** This value will be used as default access request duration for all datasets submitted by the organisation. However, there will be the opportunity to overwrite his value for each dataset.	accessRequestDuration LESS 1 WEEK: Access request process typically processed in less than a week 1-2 WEEKS: Access request process typically processed in one to two weeks 2-4 WEEKS: Access request process typically processed in two to four weeks 1-2 MONTHS: Access request process typically processed in one to two months 2-6 MONTHS: Access request process typically processed in two to six months MORE 6 MONTHS: Access request process typically processed in more than six months VARIABLE: Access request lead time is variable NOT APPLICABLE: Access request process duration is not applicable OTHER: If the typical timeframe does not fit into the broad ranges i.e. lightweight application vs linked data application, please choose "Other" and indicate the typical timeframe within the description for the dataset.
Label	Organisation Access Request Duration	
Priority	High	
Completion	Required	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	Indication of the typical duration of a data access request. Source: Schema.org Offer:deliveryLeadTime	
As A [user]	Researcher	
I Want To	Know how much time it will take to process my data access request	
So That	I can plan and decide accordingly	
Survey: Metadata Needed? (5 = Definitely Needed)	3.75/5	
Survey: Burden (5 = Easy)	3/5	

Access Rights	The URL of a webpage where the data access request process and/or guidance is provided. If there is more than one access process i.e. industry vs academic please provide both. If such a resource or the underlying process doesn't exist, please provide "In Progress", until both the process and the documentation are ready. This will then be used as default access rights for all datasets submitted by the organisation. However, there will be the opportunity to overwrite his value for each dataset.	url //Regular expression (regex) to validate URL: x ==~ /((([A-Za-z]{3,9}:(?:\/\/)?)(?:[-;:&=\+\$,\w]+@)?[A-Za-z0-9.-]+(:[0-9]+)?|(?:www.|[-;:&=\+\$,\w]+@)[A-Za-z0-9.-]+)((?:\/[\+~%\/.\w-_]*)?\??(?:[-\+=&;%@.\w_]*)#?(?:[\w]*))?)/
Title	Organisation Access Rights	
Priority	High	

Completion	Mandatory
Cardinality	Min Occurs: 1 Max Occurs: *
Definition	Information about who can access the resource or an indication of its security status. Usage Note: Access Rights MAY include information regarding access or restrictions based on privacy, security, or other policies. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dct:access_rights Range: dct:RightsStatement Source: dcat:distribution_access_rights
As A [user]	Researcher
I Want To	Know what I can do with the data
So That	I can ensure that I have a legal and ethical basis for analysing it in my research
Schema.org alternative	Dataset/conditionsOfAccess
Notes for HOP	At registration organisations will have the opportunity to add a default access rights for their organisation. If users would like to change default values, they will be given the opportunity to do so by 'un-checking' a tick box next to the item. If the defaults are updated, then all datasets that have not been 'un-checked' will be overwritten with the new value.
Survey: Metadata Needed? (5 = Definitely Needed)	4.5/5
Survey: Burden (5 = Easy)	4.14/5

Access Service	Please provide a brief description of the data access services that are available including:
	• environment that is currently available to researchers
	• additional consultancy and services
	• any indication of costs associated
	
	If no environment is currently available, please indicate the current plans and timelines when and how data will be made available to researchers
	
	Note: This value will be used as default access environment for all datasets submitted by the organisation. However, there will be the opportunity to overwrite this value for each dataset.
		longDescription
		
		//Regular expression to validate min, max character limit, min 2, max 5000
		X==~ /.{2,5000}/

Title	Organisation Access Service	
Priority	High	
Completion	Required	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	Indication of the data access environment that will be made available to researchers	
Vocabulary: DCAT	dcat:accessService	
As A [user]	Researcher	
I Want To	Know how I can access the data	
So That	I can plan and decide accordingly	

Survey: Metadata Needed? (5 = Definitely Needed)	NA
Survey: Burden (5 = Easy)	NA

Access Request Cost	Please provide link(s) to a webpage detailing the commercial model for processing data access requests for the organisation (if available) **Definition:** Indication of commercial model or cost (in GBP) for processing each data access request by the data custodian .	xs:string	Recommended Min Occurs: 0 Max Occurs: 1

As A [user]	Researcher
I Want To	Know how much it will cost to process my data access request
So That	So that I can budget and decide accordingly
Priority	Medium
Completion	Recommended
Cardinality	Min Occurs: 0 Max Occurs: 1
Definition	Indication of cost (in GBP) for processing each data access request by the data custodian.
As A [user]	Researcher
I Want To	Know how much it will cost to process my data access request
So That	So that I can budget and decide accordingly
Survey: Metadata Needed? (5 = Definitely Needed)	3.88/5
Survey: Burden (5 = Easy)	2.86/5

Data Use Limitation	Please provide an indication of consent permissions for datasets and/or materials, and relates to the purposes for which datasets and/or material might be removed, stored or used. **Notes**: where there are existing data-sharing arrangements such as the HDR UK HUB data sharing agreement or the NIHR HIC data sharing agreement this should be indicated within access rights. This value will be used as terms for all datasets submitted by the organisation. However, there will be the opportunity to overwrite this value for each dataset.	dataUseLimitation

GENERAL RESEARCH USE	This data use limitation indicates that use is allowed for general research use for any research purpose
GENETIC STUDIES ONLY	This data use limitation indicates that use is limited to genetic studies only (i.e., no phenotype-only research)
NO GENERAL METHODS RESEARCH	This data use limitation indicates that use includes methods development research(e.g., development of software or algorithms) only within the bounds of other use limitations.
NO RESTRICTION	This data use limitation indicates there is no restriction on use.
RESEARCH SPECIFIC RESTRICTIONS	This data use limitation indicates that use is limited to studies of a certain research type.
RESEARCH USE ONLY	This data use limitation indicates that use is limited to research purposes (e.g., does not include its use in clinical care).

		NO LINKAGE	This data use limitation indicates there is a restriction on linking to any other datasets
Title	Organisation Data Use Limitation		
Priority	High		
Completion	Recommended		
Cardinality	Min Occurs: 0 Max Occurs: *		
As A [user]	Researcher		
I Want To	Know the license that applies to the data		
So That	I Know the legal information that describes how the distribution is made available		
Survey: Metadata Needed? (5 = Definitely Needed)	4.13/5 (consulted on previous term - license)		
Survey: Burden (5 = Easy)	3.14 (consulted on previous term - license)		
GA4GH	https://www.ebi.ac.uk/ols/ontologies/duo/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FDUO_0000001		
GAG4H Definition	A data item that is used to indicate consent permissions for datasets and/or materials, and relates to the purposes for which datasets and/or material might be removed, stored or used.		

Data Use Requirements	Please indicate if there are any additional conditions set for use if any. Please ensure that these restrictions are documented in access rights information. Note: this value will be used as terms for all datasets submitted by the organisation. However, there will be an opportunity to overwrite this value for each dataset.	dataUseRequirements	
		COLLABORATION REQUIRED	This requirement indicates that the requestor must agree to collaboration with the primary study investigator(s).
		ETHICS APPROVAL REQUIRED	This requirement indicates that the requestor must provide documentation of local IRB/ERB approval.
		GEOGRAPHICAL RESTRICTIONS	This requirement indicates that use is limited to within a specific geographic region.
		INSTITUTION SPECIFIC RESTRICTIONS	This requirement indicates that use is limited to use within an approved institution.
		NOT FOR PROFIT USE	This requirement indicates that use of the data is limited to not-for-profit organizations and not-for-profit use, non-commercial use.
		PROJECT SPECIFIC RESTRICTIONS	This requirement indicates that use is limited to use within an approved project.
		PUBLICATION MORATORIUM	This requirement indicates that requestor agrees not to publish results of studies until a specific date
		PUBLICATION REQUIRED	This requirement indicates that requestor agrees to make results of studies using the data available to the larger scientific community.
		RETURN TO DATABASE OR RESOURCE	This requirement indicates that the requestor must return derived/enriched data to the database/resource.
		TIME LIMIT ON USE	This requirement indicates that use is approved for a specific number of months.
		USER SPECIFIC RESTRICTION	This requirement indicates that use is limited to use by approved users.
Title	Organisation Data Use Requirements		
Default	If such a legal basis does not exist e.g. the data set is being published for the first time, then please provide “In Progress”, until such time the legal terms and conditions are ready.		
Priority	High		

Completion	Recommended
Cardinality	Min Occurs: 0 Max Occurs: *
As A [user]	Researcher
I Want To	Know the license that applies to the data
So That	I know the legal information that describes how the distribution is made available
Survey: Metadata Needed? (5 = Definitely Needed)	4.13/5 (consulted on previous term - license)
Survey: Burden (5 = Easy)	3.14 (consulted on previous term - license)
GA4GH	https://www.ebi.ac.uk/ols/ontologies/duo/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FDUO_0000001
GAG4H Definition	A data item that is used to indicate consent permissions for datasets and/or materials, and relates to the purposes for which datasets and/or material might be removed, stored or used.

Documentation

Documentation can include a rich text description of the dataset or links to media such as documents, images, presentations, videos or links to data dictionaries, profiles or dashboards. Organisations are required to confirm that they have permission to distribute any additional media.

Description	An HTML account of the data that provides context and scope of the data, limited to 3000 characters (1 Page), and/or a resolvable URL that describes the dataset. Additional information can be recorded and included using media.
	
	Notes: A WYSIWYG editor will be available through the onboarding portal.
		description
		
		//Regular expression to validate min, max character limit : min 2, max 3000
		X==~ /.{2,3000}/

Priority	High	
Completion	Required	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	A free-text account of the record. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dc:description Range: rdfs:Literal Source: dcat:record_description See Schema.org Dataset:description	
As A [user]	Researcher	
I Want To	Know the description	
So That	I can understand information about the dataset	
HOP Notes	A WYSIWYG editor will be available through the onboarding portal.	

Survey: Metadata Needed? (5 = Definitely Needed)	4.5/5
Survey: Burden (5 = Easy)	4/5

Associated Media	Please provide any media associated with the Gateway Organisation using a valid URI for the content. This is an opportunity to provide additional context that could be useful for researchers wanting to understand more about the dataset and its relevance to their research question. The following formats will be accepted .jpg, .png or .svg, .pdf, .xslx or .docx.
	
	Note: media asset can be hosted by the organisation or uploaded using the onboarding portal.
		url
		
		//Regular expression (regex) to validate URL:
		x==~ /https?:\/\/(www\.)?[-a-zA-Z0-9@:%._\+~#=]{1,256}\.[a-zA-Z0-9()]{1,6}\b([-a-zA-Z0-9()@:%_\+.~#?&//=]*)/

Example	This could be a PDF Document that describes methodology or further detail about the datasets, or a graph or charts that provides further context about the dataset.	
Priority	Low	
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: *	
Definition	A media object that encodes this CreativeWork. This property is a synonym for encoding. Vocabulary: https://schema.org/associatedMedia	
As A [user]	Researcher	
I Want To	Know the description	
So That	I can understand information about the dataset	
HOP Notes	Users must confirm that they have permission to publish any media	
Survey: Metadata Needed? (5 = Definitely Needed)	NA	
Survey: Burden (5 = Easy)	NA	

Is Part Of	Please complete only if the dataset is part of a group or family of datasets i.e. Hospital Episode Statistics has several constituents. **Note:** Schema will support url, 80 characters or "NOT APPLICABLE"	xs:string
Examples	Hospital Episodes Statistics datasets (A&E, APC, OP, AC MSDS).	
Default	If no such group exists please provide “Not Applicable”.	
Priority	Medium	
Completion	Required	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	Indicates an item or CreativeWork that this item, or CreativeWork (in some sense), is part of. Inverse-property: hasPart.	
As A [user]	Researcher	

I Want To	Know whether the dataset is part of a wider group
So That	I can find other datasets of the group
HOP Notes	Default of " Not Applicable " will be provided, unless edited by the user.
Survey: Metadata Needed? (5 = Definitely Needed)	3.43/5
Survey: Burden (5 = Easy)	3.57/5

Coverage

This information includes attributes for geographical and temporal coverage, cohort details etc. to enable a deeper understanding of the dataset content so that researchers can make decisions about the relevance of the underlying data.

Spatial	The geographical area covered by the dataset. **Notes:** It is recommended that links are to entries in a well-maintained gazetteer such as https://www.geonames.org/	url //Regular expression (regex) to validate URL: x==~ /https?:\/\/(www\.)?[-a-zA-Z0-9@:%._\+~#=]{1,256}\.[a-zA-Z0-9()]{1,6}\b([-a-zA-Z0-9()@:%_\+.~#?&//=]*)/
Example	https://www.geonames.org/11609043/north-yorkshire.html	
Label	Geographic Coverage	
Priority	Medium	
Completion	Recommended	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	Usage Note: The spatial coverage of a dataset may be encoded as an instance of dct:Location or may be indicated using a URI reference (link) to a resource describing a location. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dct:spatial Range: dct:Location (A spatial region or named place) Source: dcat:dataset_spatial See also: Schema.org Dataset/spatialCoverage	
Notes	https://what3words.com/daring.lion.race	
As A [user]	Researcher	
I Want To	Know the geographic are covered by the dataset	
So That	I can contextualise my research	
Survey: Metadata Needed? (5 = Definitely Needed)	4/5	
Survey: Burden (5 = Easy)	3.86/5	

Typical Age Range	**Guidance:** Please indicate the age range in whole years of participants in the dataset. Please provide range in the following format '[min age] – [max age]' where both the minimum and maximum are whole numbers (integers).	ageRange //Regular Expression (regex) to validate age range according to format: x ==~ /(150|1[0-4][0-9]|[0-9]|[1-8][0-9]|9[0-9])-(150|1[0-4][0-9]|[0-9]|[1-8][0-9]|9[0-9])/
Example	'0-18', '11-90'	
Label	Age Range	
Priority	High	
Completion	Recommended	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	The typical age range of participants in the dataset Vocabulary: Schema.org https://schema.org/typicalAgeRange	
As A [user]	Researcher	
I Want To	Know the age range of individuals included within data	
So That	I can understand whether data is suitable for my needs	
Gateway Notes	Gateway will derive filters based on grouping the information provided however, the onboarding portal will collect the granular information, if this is available.	
Survey: Metadata Needed? (5 = Definitely Needed)	3.25/5	
Survey: Burden (5 = Easy)	3.71/5	

Physical Sample Availability	Availability of physical samples associated with the dataset. If samples are available, please indicate the types of samples that are available. More than one type may be provided. If sample are not yet available, please provide "AVAILABILITY TO BE CONFIRMED". If samples are not available, then please provide "NOT AVAILABLE".	physicalSampleAvailability

NOT AVAILABLE	Samples associated with the dataset are not available
BONE MARROW	Bone marrow samples associated with the data are available
CANCER CELL LINES	Cancer cell line samples associated with the data are available
CORE BIOPSY	Core biopsy samples associated with the data are available
CDNA/MRNA	CDNA/MRNA samples associated with the data are available
DNA	DNA samples associated with the data are available
FAECES	Faeces samples associated with the data are available
IMMORTALIZED CELL LINES	Immortalized cell line samples associated with the data are available
MICRORNA	MicroRNA samples associated with the data are available
PERIPHERAL BLOOD CELLS	Peripheral blood cell samples associated with the data are available
PLASMA	Plasma samples associated with the data are available

PM TISSUE	PM Tissue samples associated with the data are available
PRIMARY CELLS	Primary cell samples associated with the data are available
RNA	RNA samples associated with the data are available
SALIVA	Saliva samples associated with the data are available
SERUM	Serum samples associated with the data are available
SWABS	Swab samples associated with the data are available
TISSUE	Tissue samples associated with the data are available
URINE	Urine samples associated with the data are available
WHOLE BLOOD	Whole blood samples associated with the data are available
AVAILABILITY TO BE CONFIRMED	Availability of samples is currently being confirmed
OTHER	Other types of sample available

Kindly provided by the Tissue Directory Coordinating Centre

Default	NOT AVAILABLE
Priority	High
Completion	Recommended
Cardinality	Min Occurs: 0 Max Occurs: *
As A [user]	Researcher
I Want To	Know if there are any samples associated with the data available
So That	I can request access to samples for my research
Survey: Metadata Needed?	3.38/5 (5 = Definitely Needed)
Survey: Burden	3.43/5 (5 = Easy)

Followup	If known, what is the typical time span that a patient appears in the dataset (follow up period).	followup

0 - 6 MONTHS	Data typically available for a patient over a 0-6 month period
6 - 12 MONTHS	Data typically available for a patient over a 6-12 month period
1 - 10 YEARS	Data typically available for a patient over a 1-10 year period
> 10 YEARS	Data typically available for a patient for over a 10 year period
UNKNOWN	Timespan is Unknown
OTHER	Data available for a patient over another time period

Pathway	Please indicate if the dataset is representative of the patient pathway and any limitations the dataset may have with respect to pathway coverage. This could include if the dataset is from a	description

	single speciality or area, a single tier of care, linked across two tiers (e.g. primary and secondary care), or an integrated care record covering the whole patient pathway.	//Regular expression to validate min, max character limit : min 2, max 3000 X==~ /.{2,3000$/

Provenance

Provenance information allows researchers to understand data within the context of its origins and can be an indicator of quality, authenticity and timeliness.

Origin

Indicators of the origin of the data.

Purpose	Please indicate the purpose(s) that the dataset was collected.	purpose STUDY: Data collected for a specific research study. DISEASE REGISTRY: Data collected as part of a disease registry. TRIAL: Data collected for as part of a clinical trial. CARE: Data collected as part of routine clinical care. AUDIT: Data collected as part of an audit programme ADMINISTRATIVE: Data collected for administrative and management information purposes FINANCIAL: Data collected either for payments or for billing OTHER: Data collected for other purpose
Priority	High	
Completion	Recommended	
Cardinality	Min Occurs: 0 Max Occurs: *	
As A [user]	Researcher	
I Want To	Know the why the data was collected	
So That	I can understand if it is relevant to my use cases	
Definition	The purpose of the ResourcePackage, why the ResourcePackage took place. This should include detailed information on the investigator's primary ResourcePackage questions or hypotheses as well as information on any legal basis for the data collection, such as laws requiring the collection of census data for apportionment purposes. Legal or other authorization should be provided in detail within AuthorizationSource. Supports multiple language versions of the same content as well as optional formatting of the content. Vocabulary: DDI: https://ddialliance.org/Specification/DDI-Lifecycle/3.3/XMLSchema/FieldLevelDocumentation/	
HOP Notes	Multiple checkbox	

Source	Please indicate the source of the data extraction.	source EPR: Data Extracted from Electronic Patient Record ELECTRONIC SURVEY: Data has been extracted from electronic surveys

LIMS	Data has been extracted from a laboratory information management system
PAPER BASED	Data has been extracted from paper forms
FREETEXT NLP	Data has been extracted from unstructured freetext using natural language processing
MACHINE GENERATED	Data has been machine generated i.e. imaging
OTHER	Data has been extracted by another bespoke process

Priority	High
Completion	Required
Cardinality	Min Occurs: 0 Max Occurs: 1
Definition	A related resource from which the described resource is derived.
As A [user]	Researcher
I Want To	Know the source of the data
So That	I can understand the quality of the data source
Vocabulary	https://dublincore.org/specifications/dublin-core/dcmi-terms/#source

Collection Situation	Pleases indicate the setting(s) where data was collected. Multiple settings may be provided	setting

CLINIC	Specific clinic such as antenatal clinic
PRIMARY CARE	General Medical Practitioner Practice
ACCIDENT AN EMERGENCY	Accident emergency department
OUTPATIENTS	Outpatient care
IN-PATIENTS	In-patient care
SERVICES	Services such as drug misuse or blood transfusion
COMMUNITY	Community settings
HOME	Home setting
PRIVATE	Private Medical Clinic
PHARMACY	Pharmacy
OTHER	Other setting

Priority	High
Completion	Required
Label	Setting
Cardinality	Min Occurs: 0 Max Occurs: 1
Definition	A brief textual description or classification of the collection situation. Supports the use of an external controlled vocabulary Vocabulary: DDI: https://ddialliance.org/Specification/DDI-Lifecycle/3.2/XMLSchema/FieldLevelDocumentation/schemas/datacollection_xsd/elements/CollectionSituation.html
As A [user]	Researcher
I Want To	Know the source of the data
So That	I can understand the quality and provenance of the data

Temporal

Information indicative of temporal provenance.

<table>
<tr><td>Accrual Periodicity</td><td>Please indicate the frequency of publishing. If a dataset is published regularly please choose a publishing periodicity from the constrained list and indicate the next release date. When the release date becomes historical, a new release date will be calculated based on the publishing periodicity.

If a dataset has been published and will remain static please indicate that it is static and indicated when it was released. If a dataset is released on an irregular basis or “on-demand” please indicate that it is Irregular and leave release date as null.

If a dataset can be published in real-time or near-real-time please indicate that it is continuous and leave release date as null. Notes: see https://www.dublincore.org/specifications/dublin-core/collection-description/frequency/</td><td>periodicity

<table><tr><td>STATIC</td><td>Dataset published once.</td></tr><tr><td>IRREGULAR</td><td>Dataset published at uneven intervals.</td></tr><tr><td>CONTINUOUS</td><td>Dataset published without interruption.</td></tr><tr><td>BIENNIAL</td><td>Dataset published every two years.</td></tr><tr><td>ANNUAL</td><td>Dataset published occurs once a year.</td></tr><tr><td>BIANNUAL</td><td>Dataset published twice a year</td></tr><tr><td>QUARTERLY</td><td>Dataset published every three months.</td></tr><tr><td>BIMONTHLY</td><td>Dataset published every two months.</td></tr><tr><td>MONTHLY</td><td>Dataset published once a month.</td></tr><tr><td>BIWEEKLY</td><td>Dataset published every two weeks.</td></tr><tr><td>WEEKLY</td><td>Dataset published once a week.</td></tr><tr><td>SEMIWEEKLY</td><td>Dataset published twice a week.</td></tr><tr><td>DAILY</td><td>Dataset published once a day.</td></tr><tr><td>OTHER</td><td>Dataset published using other interval</td></tr></table></td></tr>
<tr><td>Priority</td><td colspan="2">High</td></tr>
<tr><td>Label</td><td colspan="2">Periodicity</td></tr>
<tr><td>Completion</td><td colspan="2">Mandatory</td></tr>
<tr><td>Cardinality</td><td colspan="2">Min Occurs: 1
Max Occurs: 1</td></tr>
<tr><td>Definition</td><td colspan="2">The frequency at which dataset is published.
Usage Note: The value of dct:accrualPeriodicity gives the rate at which the dataset-as-a-whole is updated. This may be complemented by dcat:temporalResolution to give the time between collected data points in a time series.
Vocabulary: Data Catalog Vocabulary (DCAT)
RDF Property: dct:accrualPeriodicity
Range: dct:Frequency (A rate at which something recurs)
Source: dcat:dataset_frequency
See also: Schema.org Dataset/temporalCoverage</td></tr>
<tr><td>As A [user]</td><td colspan="2">Researcher</td></tr>
<tr><td>I Want To</td><td colspan="2">Know the frequency of the dataset update</td></tr>
<tr><td>So That</td><td colspan="2">I know the timeliness of the data</td></tr>
<tr><td>HOP Notes</td><td colspan="2">For periodic releases we should notify the user to automatically create a new version of the dataset.</td></tr>
<tr><td>Survey: Metadata Needed? (5 = Definitely Needed)</td><td colspan="2">4.25/5</td></tr>
</table>

Survey: Burden (5 = Easy)	4.29/5

Distribution Release Date	Date of the latest release of the dataset. If this is a regular release i.e. quarterly, or this is a static dataset please complete this alongside Periodicity. If this is Irregular or Continuously released please leave this blank. **Notes**: Periodicity and release date will be used to determine when the next release is expected. i.e. the next release date can then be calculated automatically. E.g. if the release date is documented as 01/01/2020 and it is now 20/04/2020 and there is a quarterly release schedule, the latest release will be calculated as 01/04/2020.	xs:dateTime

Priority	High
Label	Release Date
Completion	Required (if applicable)
Cardinality	Min Occurs: 0 Max Occurs: 1
Definition	Date and Time compliant string [DATETIME] and typed using the appropriate XML Schema datatype [XMLSCHEMA11-2] (xsd:gYear , xsd:gYearMonth , xsd:date , or xsd:dateTime). Usage Note: This property SHOULD be set using the first Known date of issuance. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dct:issued Range: rdfs:Literal Source: dcat:distribution_release_date See Schema.org Dataset/datePublished
As A [user]	Researcher
I Want To	Know the publication date
So That	I Know how recent, relevant and timely the dataset is
HOP Notes	Validation should be included to ensure that release date is completed if periodicity is regular
Survey: Metadata Needed? (5 = Definitely Needed)	4/5
Survey: Burden (5 = Easy)	3.43/5

Start Date	The start of the time period that the dataset provides coverage for. If there are multiple cohorts in the dataset with varying start dates, please provide the earliest date and use the description or the media attribute to provide more information.	xs:dateTime

Priority	High
Completion	Mandatory
Cardinality	Min Occurs: 1 Max Occurs: 1
Definition	The start of the period. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dcat:startDate Domain: dcat:PeriodOfTime Range: rdfs:Literal encoded using the relevant ISO 8601 Date and Time compliant string [DATETIME] and typed using the appropriate XML Schema datatype [XMLSCHEMA11-2] (xsd:gYear , xsd:gYearMonth , xsd:date , or xsd:dateTime). Source: dcat:period_start_date See also: Schema.org Dataset/temporalCoverage
As A [user]	Researcher
I Want To	Know the time period from when data is available
So That	I can use data that is within the time period I am interested and understand if there is enough history
Survey: Metadata Needed? (5 = Definitely Needed)	4.13/5
Survey: Burden (5 = Easy)	4.14/5

End Date	The end of the time period that the dataset provides coverage for. If the dataset is "Continuous" and has no known end date, please leave blank. If there are multiple cohorts in the dataset with varying end dates, please provide the latest date.	endDate // value is dateTime or date or NOT APPLICABLE x ==~ /(\d{4}-[01]\d-[0-3]\dT[0-2]\d:[0-5]\d:[0-5]\d(?:\.\d+)?Z?)|(\d{4}-[01]\d-[0-3]\d)|(NOT APPLICABLE)/

Priority	High
Completion	Required (if applicable)
Cardinality	Min Occurs: 0 Max Occurs: 1
Definition	The end of the period. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dcat:endDate Domain: dcat:PeriodOfTime Source: dcat:period_end_date Range: rdfs:Literal encoded using the relevant ISO 8601 Date and Time compliant string [DATETIME] and typed using the appropriate XML Schema datatype [XMLSCHEMA11-2] (xsd:gYear , xsd:gYearMonth , xsd:date , or xsd:dateTime). Source : Schema.org Dataset/temporalCoverage
As A [user]	Researcher
I Want To	Know the time period until when data is available
So That	I can use data that is within the time period I am interested in and understand if there is enough history.

Survey: Metadata Needed? (5 = Definitely Needed)	4.13/5
Survey: Burden (5 = Easy)	3.86/5

Time Lag	Please indicate the typical time-lag between an event and the data for that event appearing in the dataset.	timeLag LESS 1 WEEK: Typical time lag of less than a week 1-2 WEEKS: Typical time lag of one to two weeks 2-4 WEEKS: Typical time lag of two to four weeks 1-2 MONTHS: Typical time lag of one to two months 2-6 MONTHS: Typical time lag of two to six months 6 MONTHS PLUS: Typical time lag of more than six months VARIABLE: Variable time lag NOT APPLICABLE: Not Applicable i.e. static dataset OTHER: Other time lag
Default	VARIABLE	
Priority	High	
Label	Periodicity	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	

Accessibility

Accessibility information allows researchers to understand access, usage, limitations, formats, standards and linkage or interoperability with toolsets.

Usage

This section includes information about how the data can be used and how it is currently being used.

Data Use Limitation	Please provide an indication of consent permissions for datasets and/or materials, and relates to the purposes for which datasets and/or material might be removed, stored or used. NOTE: we have extended the DUO to include a value for NO LINKAGE	dataUseLimitation GENERAL RESEARCH USE: This data use limitation indicates that use is allowed for general research use for any research purpose GENETIC STUDIES ONLY: This data use limitation indicates that use is limited to genetic studies only (i.e., no phenotype-only research) NO GENERAL METHODS RESEARCH: This data use limitation indicates that use includes methods development research(e.g., development of software or algorithms) only within the bounds of other use limitations. NO RESTRICTION: This data use limitation indicates there is no restriction on use. RESEARCH SPECIFIC RESTRICTIONS: This data use limitation indicates that use is limited to studies of a certain research type. RESEARCH USE ONLY: This data use limitation indicates that use is limited to research purposes (e.g., does not include its use in clinical care). NO LINKAGE: This data use limitation indicates there is a restriction on linking to any other datasets
Title	Organisation Data Use Limitation	

Priority	High
Completion	Mandatory
Cardinality	Min Occurs: 0 Max Occurs: 1
As A [user]	Researcher
I Want To	Know the license that applies to the data
So That	I Know the legal information that describes how the distribution is made available
Survey: Metadata Needed? (5 = Definitely Needed)	4.13/5 (consulted on previous term - license)
Survey: Burden (5 = Easy)	3.14 (consulted on previous term - license)
GA4GH	https://www.ebi.ac.uk/ols/ontologies/duo/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FDUO_0000001
GAG4H Definition	A data item that is used to indicate consent permissions for datasets and/or materials, and relates to the purposes for which datasets and/or material might be removed, stored or used.

Data Use Requirements	Please indicate if there are any additional conditions set for use if any, multiple requirements may be provided. Please ensure that these restrictions are documented in access rights information.	datauseRequirements COLLABORATION REQUIRED: This requirement indicates that the requestor must agree to collaboration with the primary study investigator(s). ETHICS APPROVAL REQUIRED: This requirement indicates that the requestor must provide documentation of local IRB/ERB approval. GEOGRAPHICAL RESTRICTIONS: This requirement indicates that use is limited to within a specific geographic region. INSTITUTION SPECIFIC RESTRICTIONS: This requirement indicates that use is limited to use within an approved institution. NOT FOR PROFIT USE: This requirement indicates that use of the data is limited to not-for-profit organizations and not-for-profit use, non-commercial use. PROJECT SPECIFIC RESTRICTIONS: This requirement indicates that use is limited to use within an approved project. PUBLICATION MORATORIUM: This requirement indicates that requestor agrees not to publish results of studies until a specific date PUBLICATION REQUIRED: This requirement indicates that requestor agrees to make results of studies using the data available to the larger scientific community. RETURN TO DATABASE OR RESOURCE: This requirement indicates that the requestor must return derived/enriched data to the database/resource. TIME LIMIT ON USE: This requirement indicates that use is approved for a specific number of months. USER SPECIFIC RESTRICTION: This requirement indicates that use is limited to use by approved users.
GA4GH	https://www.ebi.ac.uk/ols/ontologies/duo/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FDUO_0000001	
GA4GH Definition	A data item that is used to indicate consent permissions for datasets and/or materials, and relates to the purposes for which datasets and/or material might be removed, stored or used.	

Resource Creator	Please provide the text that you would like included as part of any citation that credits this dataset. This is typically just the name of the publisher. No employee details should be provided.	shortDescription // value is between 2 and 1000 characters x ==~ /^.{2,1000}$/
Label	Citation Requirements	
Example	National Services Scotland	
Default	Name of the publisher	
Priority	Medium	
Completion	Mandatory	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	The entity responsible for producing the resource. Usage Note: Resources of type foaf:Agent are recommended as values for this property. See also: § 6.11 Class: Organization/Person Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dct:creator Range: foaf:term_Agent Source: dcat:resource_creator	
As A [user]	Researcher	
I want to	Know the creator	
So That	I can understand the context and means for attribution	
Survey: Metadata Needed? (5 = Definitely Needed)	3.38/5	
Survey: Burden (5 = Easy)	3.57/5	

Investigations	Please provide a link to any active projects that are using the dataset.	url //Regular expression (regex) to validate URL: x==~ /(http(s)?:\/\/.)?(www\.)?[-a-zA-Z0-9@:%._\+~#=]{2,256}\.[a-z]{2,6}\b([-a-zA-Z0-9@:%_\+.~#?&//=]*) |(In Progress)/
Cardinality	MinOccurs: 0 Max Occurs: *	

Is Referenced By	Please provide the keystone paper associated with the dataset. Also include a list of known citations, if available and should be links to existing resources where the dataset has been used or referenced. Please provide multiple entries, or if you are using a csv upload please provide them as a tab separated list. [Author] ([Year]) [Title] . [DOI] . [Source]	doi // value is doi x ==~ /10.\d{4,9}\/[\-\._\;\(\)\/\:a-z0-9A-Z]+/
Label	Citations	
Example	de Lusignan S, Correa A, Smith GE, Yonova I, Pebody R, Ferreira F, Elliot AJ, Fleming DM. RCGP Research and Surveillance Centre: 50 years' surveillance of influenza, infections, and	

	respiratory conditions. BJGP 2017; 67 (663): 440-441. DOI: https://doi.org/10.3399/bjgp17X692645.
Priority	Medium
Completion	Recommended (if applicable)
Cardinality	Min Occurs: 0 Max Occurs: *
Definition	A related resource, such as a publication, that references, cites, or otherwise points to the catalogued resource. Usage Note: In relation to the use case of data citation, when the catalogued resource is a dataset, the dct:isReferencedBy property allows to relate the dataset to the resources (such as scholarly publications) that cite or point to the dataset. Multiple dct:isReferencedBy properties can be used to indicate the dataset has been referenced by multiple publications, or other resources. Usage Note: This property is used to associate a resource with the resource (of type dcat:Resource) in question. For other relations to resources not covered with this property, the more generic property dcat:qualifiedRelation can be used. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dct:isReferencedBy Source: dcat:resource_is_referenced_by
As A [user]	Researcher
I Want To	Know how to cite the asset
So That	I can appropriately cite the resource in my research publications
Survey: Metadata Needed? (5 = Definitely Needed)	3.38/5
Survey: Burden (5 = Easy)	2.71/5

Access

This section includes information about data access.

Access Rights	The URL of a webpage where the data access request process and/or guidance is provided. If there is more than one access process i.e. industry vs academic please provide both.
	
	If such a resource or the underlying process doesn't exist, please provide "In Progress", until both the process and the documentation are ready.
		url
		
		//Regular expression (regex) to validate URL:
		x==~ /(http(s)?:\/\/.)?(www\.)?[-a-zA-Z0-9@:%._\+~#=]{2,256}\.[a-z]{2,6}\b([-a-zA-Z0-9@:%_\+.~#?&//=]*) |(In Progress)/

Default	Organisation Access Rights	
Priority	High	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: *	
Definition	Information about who can access the resource or an indication of its security status. Usage Note: Access Rights MAY include information regarding access or restrictions based on privacy, security, or other policies. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dct:access_rights Range: dct:RightsStatement	

	Source: dcat:distribution_access_rights
As A [user]	Researcher
I Want To	Know what I can do with the data
So That	I can ensure that I have a legal and ethical basis for analysing it in my research
Schema.org alternative	Dataset/conditionsOfAccess
Notes for HOP	At registration organisations will have the opportunity to add a default access rights for their organisation. If users would like to change default values, they will be given the opportunity to do so by 'un-checking' a tick box next to the item. If the defaults are updated, then all datasets that have not been 'un-checked' will be overwritten with the new value.
Survey: Metadata Needed? (5 = Definitely Needed)	4.5/5
Survey: Burden (5 = Easy)	4.14/5

Access Service	Please provide a brief description of the data access services that are available including: • environment that is currently available to researchers • additional consultancy and services • any indication of costs associated If no environment is currently available, please indicate the current plans and timelines when and how data will be made available to researchers **Note:** This value will be used as default access environment for all datasets submitted by the organisation. However, there will be the opportunity to overwrite this value for each dataset.	longDescription //Regular expression to validate min, max character limit, min 2, max 5000 X==~ /.{2,5000}/
Label	Research Environment, Service, Cost	
Priority	High	
Completion	Required	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	A data service that gives access to the distribution of the dataset Vocabulary: Data Catalog Vocabulary (DCAT RDF Property: dcat:accessService Usage Notes: cat:accessService SHOULD be used to link to a description of a dcat:DataService that can provide access to this distribution.	
As A [user]	Researcher	
I Want To	Know how I can access the data	
So That	I can plan and decide accordingly	
Survey: Metadata Needed? (5 = Definitely Needed)	NA	
Survey: Burden (5 = Easy)	NA	

Access Request Cost	Please provide link(s) to a webpage detailing the commercial model for processing data access requests for the organisation (if available) **Definition:** Indication of commercial model or cost (in GBP) for processing each data access request by the data custodian .	xs:string	Recommended Min Occurs: 0 Max Occurs: 1
As A [user]	Researcher		
I Want To	Know how much it will cost to process my data access request		
So That	So that I can budget and decide accordingly		
Priority	Medium		
Completion	Recommended		
Cardinality	Min Occurs: 0 Max Occurs: 1		
Definition	Indication of cost (in GBP) for processing each data access request by the data custodian.		
As A [user]	Researcher		
I Want To	Know how much it will cost to process my data access request		
So That	So that I can budget and decide accordingly		
Survey: Metadata Needed? (5 = Definitely Needed)	3.88/5		
Survey: Burden (5 = Easy)	2.86/5		

Delivery Lead Time	Please provide an indication of the typical processing times based on the types of requests typically received.	accessRequestDuration LESS 1 WEEK: Access request process typically processed in less than a week 1-2 WEEKS: Access request process typically processed in one to two weeks 2-4 WEEKS: Access request process typically processed in two to four weeks 1-2 MONTHS: Access request process typically processed in one to two months 2-6 MONTHS: Access request process typically processed in two to six months MORE 6 MONTHS: Access request process typically processed in more than six months VARIABLE: Access request lead time is variable NOT APPLICABLE: Access request process duration is not applicable OTHER: If the typical timeframe does not fit into the broad ranges i.e. lightweight application vs linked data application, please choose "Other" and indicate the typical timeframe within the description for the dataset.
Label	Access Request Duration	
Priority	High	
Completion	Required	
Cardinality	Min Occurs: 0 Max Occurs: 1	

Definition	Indication of the typical duration of a data access request. Source: https://schema.org/deliveryLeadTime
As A [user]	Researcher
I Want To	Know how much time it will take to process my data access request
So That	I can plan and decide accordingly
Survey: Metadata Needed? (5 = Definitely Needed)	3.75/5
Survey: Burden (5 = Easy)	3/5

Jurisdiction	Please use country code from ISO 3166-1 country codes and the associated ISO 3166-2 for regions, cities, states etc. for the country/state under whose laws the data subjects' data is collected, processed and stored. Multiple Jurisdictions may be provided (if applicable).	isocountryCode // value is decimal number x ==~ /^[A-Z]{2}(-[A-Z]{2,3})?$/ Microsoft Excel 97-2004
Default	GB-ENG	
Priority	High	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: *	
Definition	A named and identified geospatial area with defined borders which is used for exercising the action of the Rule. An IRI MUST be used to represent this value. Note: A code value for the area and source of the code must be presented in the Right Operand. Example: the [iso3166] Country Codes or the Getty Thesaurus of Geographic Names. Narrower terms: spatialCoordinates. Vocabulary: Dublin Core Source: http://purl.org/dc/terms/Jurisdiction	
Notes for HOP	Multi-select needed	
As A [user]	Researcher	
I Want To	Know about potential constraints on the data due to jurisdiction	
So That	I can tailor my request and research accordingly (or realise the data isn't suitable)	
Survey: Metadata Needed? (5 = Definitely Needed)	3.5/5	
Survey: Burden (5 = Easy)	3.71/5	

Data Controller	Data Controller means a person/entity who (either alone or jointly or in common with other persons/entities) determines the purposes for which and the way any Data Subject data, specifically personal data or are to be processed.
	
	Notes: For most organisations this will be the same as the publisher of the dataset. If this is not the case, please indicate that there is a different controller.
	
	If there is a different controller please complete the Data Processor attribute to indicate that the publisher is a Processor rather than the data controller.
		longDescription
		
		//Regular expression to validate min, max character limit : min 2, max 5000
		X==~ /^.{2,5000}$/

Examples	In some cases, there may be multiple data controllers i.e. GP data. If this is the case please indicate the fact in a free-text field and describe the data sharing arrangement or a link to it, so that this can be understood by research users.	
Priority	High	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Definition	Source: ICO Key Terms: Controllers and Processors , Article 4.7 of GDPR Vocabulary: Data Privacy Vocabulary Source: dpv:DataController	
As A [user]	Processor	
I Want To	Know who has ultimate responsibility for the dataset (if not the publisher)	
So That	I can continue to process their data	
Survey: Metadata Needed? (5 = Definitely Needed)	3.75/5	
Survey: Burden (5 = Easy)	3.86/5	

Data Processor	A Data Processor, in relation to any Data Subject data, specifically personal data, means any person/entity (other than an employee of the data controller) who processes the data on behalf of the data controller.
	
	Notes: Required to complete if the Publisher is the Data Processor rather than the data controller.
		longDescription
		
		//Regular expression to validate min, max character limit : min 2, max 5000
		X==~ /^.{2,5000}$/

Default	If the Publisher is also the Data Controller please provide “Not Applicable”.	
Priority	High	
Completion	Required (if applicable)	
Cardinality	Min Occurs: 0 Max Occurs:1	
Definition	Source: ICO Key Terms: Controllers and Processors , Article 4.7 of GDPR Vocabulary: Data Privacy Vocabulary Source: dpv:DataProcessor	
As A [user]	Administrator	
I Want To	Know who the data processor is	
So That	I that can understand the legal basis for processing	

Survey: Metadata Needed? (5 = Definitely Needed)	3.75/5
Survey: Burden (5 = Easy)	3.43/5

Formats & Standards

Section includes technical attributes for language vocabularies, sizes etc. and gives researchers facts about and processing the underlying data in the dataset.

		controlledVocabulary	
Vocabulary Encoding Scheme	List any relevant terminologies / ontologies / controlled vocabularies, such as ICD 10 Codes, NHS Data Dictionary National Codes or SNOMED CT International, that are being used by the dataset. If the controlled vocabularies are local standards, please make that explicit. If you are using a standard that has not been included in the list, please use "other" and contact support desk to ask for an addition. Notes: More than one vocabulary may be provided.	**LOCAL**	Local Coding Standard
		OPCS4	https://www.datadictionary.nhs.uk/web_site_content/supporting_information/clinical_coding/opcs_classification_of_interventions_and_procedures.asp
		READ	https://digital.nhs.uk/services/terminology-and-classifications/read-codes
		SNOMED CT	http://www.snomed.org/
		SNOMED RT	https://confluence.ihtsdotools.org/display/DOCGLOSS/SNOMED+RT
		DM+D	https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/scci0052-dictionary-of-medicines-and-devices-dm-d
		NHS NATIONAL CODES	https://www.datadictionary.nhs.uk/
		NHS SCOTLAND NATIONAL CODES	https://www.ndc.scot.nhs.uk/Data-Dictionary/
		NHS WALES NATIONAL CODES	http://www.datadictionary.wales.nhs.uk/
		ODS	https://digital.nhs.uk/services/organisation-data-service
		LOINC	https://loinc.org/
		ICD10	https://www.who.int/classifications/icd/icdonlineversions/en/
		ICD10CM	https://www.cdc.gov/nchs/icd/icd10cm.htm
		ICD10PCS	https://ec.europa.eu/eip/ageing/standards/healthcare/e-health/icd-10-pcs_en
		ICD9CM	https://www.cdc.gov/nchs/icd/icd9cm.htm
		ICD9	https://www.cdc.gov/nchs/icd/icd9.htm

ICDO3	https://www.who.int/classifications/icd/adaptations/oncology/en/
AMT	https://www.digitalhealth.gov.au/about-the-agency/tenders-and-offers/community-pharmacy-software-industry-partnership-offer/Webinar%20-%20Australian%20Medicines%20Terminology%20(AMT)%20and%20Implementation%20Options%2001032017.pdf
APC	https://www.acep.org/administration/reimbursement/reimbursement-faqs/apc-ambulatory-payment-classifications-faq/
ATC	https://www.whocc.no/atc_ddd_index/
CIEL	https://github.com/OpenConceptLab/ocl_web/wiki/CIEL
HPO	https://hpo.jax.org/app/
CPT4	https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/Downloads/SubjecttoCLIA.pdf
DPD	https://health-products.canada.ca/dpd-bdpp/index-eng.jsp
DRG	http://www.euro.who.int/__data/assets/pdf_file/0004/162265/e96538.pdf
HEMONC	https://hemonc.org/wiki/Main_Page
JMDC	https://www.jmdc.co.jp/en/
KCD7	https://forums.ohdsi.org/t/adding-kcd7-code-korean-icd-10-to-the-omop-vocabulary/7576
MULTUM	https://www.cerner.com/solutions/drug-database
NAACCR	https://www.naaccr.org/
NDC	https://www.fda.gov/drugs/drug-approvals-and-databases/national-drug-code-directory
NDFRT	https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/NDFRT/index.html
OXMIS	https://oxrisk.com/oxmis/
RXNORM	https://www.nlm.nih.gov/research/umls/rxnorm/index.html
RXNORM EXTENSION	https://www.nlm.nih.gov/research/umls/rxnorm/index.html
SPL	https://www.fda.gov/industry/fda-resources-data-standards/structured-product-labeling-resources
OTHER	Please indicate if there is another standard that you are using. This will trigger a support ticket where you can request the addition of the terminology to the HOP.

Label	Controlled Vocabulary
Default	LOCAL
Priority	Medium
Completion	Mandatory
Cardinality	Min Occurs: 1 Max Occurs: *
Definition	An enumerated set of resources. Vocabulary: Dublin Core https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#http://purl.org/dc/dcam/VocabularyEncodingScheme
As A [user]	Researcher
I Want To	Know which vocabulary does the dataset use or provide
So That	I can use it to link to other datasets that use the same vocabularies
HOP Notes	If Other selected this should trigger a support ticket
Survey: Metadata Needed? (5 = Definitely Needed)	3.75/5
Survey: Burden (5 = Easy)	3/5

		standardisedDataModels
Conforms To	List standardised data models that the **dataset** has been **stored in** or **transformed to,** such as OMOP or FHIR. If the data is only available in a local format, please make that explicit. If you are using a standard that has not been included in the list, please use "other" and contact support desk to ask for an addition.	

HL7 FHIR	https://www.hl7.org/fhir/
HL7 V2	https://www.hl7.org/implement/standards/product_section.cfm?section=13
HL7 CDA	https://www.hl7.org/implement/standards/product_section.cfm?section=10
HL7 CCOW	https://www.hl7.org/implement/standards/product_section.cfm?section=16
LOINC	https://loinc.org/
DICOM	https://www.dicomstandard.org/
I2B2	https://www.i2b2.org/
IHE	https://www.ihe.net/resources/profiles/
OMOP	https://www.ohdsi.org/data-standardization/the-common-data-model/
OPENEHR	https://www.openehr.org/
SENTINEL	https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model
PCORNET	https://pcornet.org/data-driven-common-model/
CDISC	https://www.cdisc.org/standards/data-exchange/odm
LOCAL	In-house developed data model
OTHER	Other standardised data model

Default	LOCAL
Priority	Medium
Completion	Mandatory
Cardinality	Min Occurs: 1 Max Occurs: 1
Definition	An established standard to which the described resource conforms. “A basis for comparison; a reference point against which other things can be evaluated." [DCTERMS]) Usage Note: This property SHOULD be used to indicate the model, schema, ontology, view or profile that the catalogued resource content conforms to.” Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dct:conformsTo Range: dct:Standard Source: dcat:resource_conforms_to
As A [user]	Researcher
I Want To	Know which standards the data set conforms to
So That	I can apply models, pipelines and analysis from
Survey: Metadata Needed? (5 = Definitely Needed)	3.13/5
Survey: Burden (5 = Easy)	2.86/5

Language	This should list all the languages in which the dataset metadata and underlying data is made available.	xs:language Microsoft Excel 97-2004
Default	en	
Priority	High	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: *	
Definition	A language of the item. This refers to the natural language used for textual metadata (i.e. titles, descriptions, etc) of a catalogued resource (i.e. dataset or service) or the textual values of a dataset distribution. Usage Note: Resources defined by the Library of Congress (ISO 639-1, ISO 639-2) SHOULD be used. If an ISO 639-1 (two-letter) code is defined for language, then its corresponding IRI SHOULD be used; if no ISO 639-1 code is defined, then IRI corresponding to the ISO 639-2 (three-letter) code SHOULD be used. Repeat this property if the resource is available in multiple languages. If representations of a dataset are available for each language separately, define an instance of dcat:Distribution for each language and describe the specific language of each distribution using dcat:language (i.e. the dataset will have multiple dcat:language values and each distribution will have just one as the value of its dcat:language property). Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dct:language	

	Range: dct:LinguisticSystem Source: dcat:language
As A [user]	Researcher
I Want To	Know the language of the dataset
So That	I understand the contents using that language
Survey: Metadata Needed? (5 = Definitely Needed)	4.25/5
Survey: Burden (5 = Easy)	4.57/5

Format	If multiple formats are available, please specify. See application, audio, image, message, model, multipart, text, video, https://www.iana.org/assignments/media-types/media-types.xhtml **Note**: If your file format is not included in the current list of formats, please indicate other. If you are using the HOP, you will be directed to a service desk page where you can request your additional format. If not, please go to: https://metadata.atlassian.net/servicedesk/customer/portal/4 to request your format.	format Application: Microsoft Excel 97-2004 Text: Microsoft Excel 97-2004 Image: Microsoft Excel 97-2004 Video: Microsoft Excel 97-2004 Audio: Microsoft Excel 97-2004
Example	text/tab-separated-values, application/sql, text/csv, image/diacom-rle	
Priority	High	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: *	
Definition	The file format of the distribution. Usage Note: dcat:mediaType SHOULD be used if the type of the distribution is defined by IANA [IANA-MEDIA-TYPES]. RDF Property: http://purl.org/dc/terms/format Range: https://dublincore.org/specifications/dublin-core/dcmi-terms/#MediaTypeOrExtent	

	Source: https://www.w3.org/TR/vocab-dcat-2/#Property:distribution_format
As A [user]	Researcher
I Want To	Know the format
So That	I Know how to process it
Survey: Metadata Needed? (5 = Definitely Needed)	3.88/5
Survey: Burden (5 = Easy)	4.57/5

Enrichment & Linkage

This section includes information about related datasets that may have previously been linked, as well as indicating if there is the opportunity to link to other datasets in the future. If a dataset has been enriched and/or derivations, scores and existing tools are available this section allows providers to indicate this to researchers.

Qualified Relation	If applicable, please provide the DOI of other datasets that have previously been linked to this dataset and their availability. If no DOI is available, please provide the title of the datasets that can be linked, where possible using the same title of a dataset previously onboarded to the HOP. **Note**: If all the datasets from Gateway organisation can be linked please indicate "ALL" and the onboarding portal will automate linkage across the datasets submitted.	linkedDatasets //Regular expression (regex) to validate URL: x ==~ /((([A-Za-z]{3,9}:(?:\/\/)?)(?:[-;:&=\+\$,\w]+@)?[A-Za-z0-9.-]+(:[0-9]+)?|(?:www.|[-;:&=\+\$,\w]+@)[A-Za-z0-9.-]+)((?:\/[\+~%\/.\w-_]*)?\??(?:[-\+=&;%@.\w_]*)#?(?:[\w]*))?)|ALL|NA/
Label	Linked Dataset	
Priority	Medium	
Completion	Recommended (if applicable)	
Cardinality	Min Occurs: 0 Max Occurs: *	
Definition	Link to a description of a relationship with another resource. Usage Note: Used to link to another resource where the nature of the relationship is Known but does not match one of the standard [DCTERMS] properties (dct:hasPart, dct:isPartOf, dct:conformsTo, dct:isFormatOf, dct:hasFormat, dct:isVersionOf, dct:hasVersion, dct:replaces, dct:isReplacedBy, dct:references, dct:isReferencedBy, dct:requires, dct:isRequiredBy) or [PROV-O] properties (prov:wasDerivedFrom, prov:wasInfluencedBy, prov:wasQuotedFrom, prov:wasRevisionOf, prov:hadPrimarySource, prov:alternateOf, prov:specializationOf). Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dcat:qualifiedRelation Sub-property of: prov:qualifiedInfluence Domain: dcat:Resource Range: dcat:Class:Relationship Source: dcat:resource_qualified_relation	
As A [user]	Researcher	
I Want To	Know if there are any datasets that have already been linked	

So That	I can use more pre-processed data in my research
Survey: Metadata Needed? (5 = Definitely Needed)	3.63/5
Survey: Burden (5 = Easy)	3.57/5

Derivation	Indicate if derived datasets or predefined extracts are available and the type of derivation available. **Notes:** Single or multiple dimensions can be provided as a derived extract alongside the dataset.	abstractText //Regular expression (regex) to validate and constrain title between 5 and 255 characters x==~ /^.{5,255}$/
Label	Derived Dataset	
Default	If such derivations are not known by the Publisher, then please provide "Not Known". If there are no derivations calculated, please indicate "Not Available".	
Examples	The Royal College of GPs derives scores such Charlson Comorbidity Score as part of their standard outputs.	
Priority	Medium	
Completion	Recommended (if applicable)	
Cardinality	Min Occurs: 0 Max Occurs: *	
Definition	A derivation is a transformation of an entity into another, an update of an entity resulting in a new one, or the construction of a new entity based on a pre-existing entity. If this is Entity:e, then it can qualify how it was derived using prov:qualifiedDerivation [a prov:Derivation; prov:entity :e; :foo :bar]. Vocabulary: PROV-O: The PROV Ontology RDF Property: prov:qualifiedDerivation Sub-property of: prov:qualifiedInfluence Domain: prov:Entity Range: prov:Derivation PROV-DM term: prov:wasDerivedFrom See also: Schema.org Dataset/isBasedOn	
As A [user]	Researcher	
I Want To	Know if there are any derivations	
So That	I can reuse pre-processed information	
Survey: Metadata Needed? (5 = Definitely Needed)	3.57/5	
Survey: Burden (5 = Easy)	3.57/5	

Tools	Please provide the URL of any analysis tools or models that have been created for this dataset and are available for further use. Multiple tools may be provided. **Note:** We encourage users to adopt a model along the lines of	url //Regular expression (regex) to validate URL:

	https://www.ga4gh.org/news/tool-registry-service-api-enabling-an-interoperable-library-of-genomics-analysis-tools/	x==~ /https?:\/\/(www\.)?[-a-zA-Z0-9@:%._\+~#=]{1,256}\.[a-zA-Z0-9()]{1,6}\b([-a-zA-Z0-9()@:%_\+.~#?&//=]*)/
Cardinality	Min Occurs: 0 Max Occurs: *	

Observations

Based on schema.org observation model and used to specify series of observations about a dataset, such as number of persons, events or findings.

We will be supporting the schema.org observation model (https://schema.org/Observation). Users will be encouraged to provide their own statistical populations as the project progresses however, initially we will encourage users to provide observations on the number of persons, number of events and number of findings included in the dataset.

Example:
Statistical Population 1
type: StatisticalPopulation
populationType: Persons
numConstraints: 0

Statistical Population 2
type: StatisticalPopulation
populationType: Events
numConstraints: 0

Statistical Population 3
type: StatisticalPopulation
populationType: Findings
numConstraints: 0

typeOf: Observation
observedNode: Statistical Population 1
measuredProperty: count
measuredValue: 32937
observationDate: “2017”

Observation

Multiple observations about the dataset may be provided and users are expected to provide at least one observation (1..*). We will be supporting the schema.org observation model (https://schema.org/Observation) with default values. Users will be encouraged to provide their own statistical populations as the project progresses.

Example:

Statistical Population 1
type: StatisticalPopulation
populationType: Persons
numConstraints: 0

Statistical Population 2
type: StatisticalPopulation
populationType: Events
numConstraints: 0

Statistical Population 3
type: StatisticalPopulation
populationType: Findings
numConstraints: 0

typeOf: Observation
observedNode: ***Statistical Population 1***
measuredProperty: count
measuredValue: 32937
observationDate: "2017"

Observed Node	Please select one of the following statistical populations for your observation.	statisticalPopulationConstrained Persons: Unique persons recorded in the dataset Events: Unique events such as procedures and prescriptions within the dataset Findings: Unique findings included in the dataset such as diagnoses, test results, images or specimens
Label	Statistical Population	
Definition	The observedNode of an Observation, often a StatisticalPopulation. Vocabulary: Schema.org https://schema.org/observedNode	
Cardinality	Min Occurs: 1 Max Occurs: 1	

Measured Value	Please provide the population size associated with the population type the dataset i.e. 1000 people in a study, or 87 images (MRI) of Knee **Usage Note:** Used with Statistical Population, which specifies the type of the population in the dataset.	xs:positiveInteger
Priority	High	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Definition	Source: Schema.org https://schema.org/measuredValue	
As A [user]	Researcher	
I Want To	Know the population size and type within data	
So That	I can understand whether data amount is suitable for my needs	
Survey: Metadata Needed? (5 = Definitely Needed)	3.25/5	

Survey: Burden (5 = Easy)	3/5

		abstractText
Disambiguating Description	If SNOMED CT term does not provide sufficient detail, please provide a description that disambiguates the population type.	//Regular expression (regex) to validate and constrain title between 5 and 255 characters x==~ /^.{5,255}$/
Priority	High	
Completion	Recommended	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	A short description of the item used to disambiguate from other, similar items. Information from other properties (in particular, name) may be necessary for the description to be useful for disambiguation. Source: Schema.org https://schema.org/disambiguatingDescription	
As A [user]	Researcher	
I Want To	Know the population size and type within data	
So That	I can understand whether data amount is suitable for my need	
Survey: Metadata Needed? (5 = Definitely Needed)	NA	
Survey: Burden (5 = Easy)	NA	

Observation Date	Please provide the date of the observation.	xs:dateTime
Vocabulary: Schema.org	https://schema.org/observationDate	
Definition	The observationDate of an Observation.	
Default to	Release Date	
Cardinality	Min Occurs: 1 Max Occurs: 1	

Measured Property	Initially this will be defaulted to "Count".	xs:string
Priority	Medium	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Vocabulary	Schema.org: The measuredProperty of an Observation, either a schema.org property, a property from other RDF-compatible systems e.g. W3C RDF Data Cube, or schema.org extensions such as GS1's. https://schema.org/measuredProperty	
As A [user]	Researcher	
I Want To	Know the population type within the data	
So That	I can understand whether data subject is suitable for my need	

HOP Notes	Initially we will want a “search” on a subset of SNOMED CT terms.
Survey: Metadata Needed? (5 = Definitely Needed)	4.13/5
Survey: Burden (5 = Easy)	4/5

Appendix 1: Change Log

HDR UK Descriptive Metadata Specifications v2				HOP Descriptive Metadata Specifications v1	
Descriptive Attribute Hierarchy			**Description**	**Level**	**Attribute Name**
Identifier			System dataset identifier.	Summary	Identifier
Version			Dataset metadata version.	Summary	Version
Revisions	URL		Revisions of Dataset metadata: Identifier.		
Revisions	Version		Revisions of Dataset metadata: Version.		
Issued			Dataset descriptive metadata creation date.		
Modified			Dataset descriptive metadata modification date.		
Summary	Title		Title of the dataset. It should provide a short description.	Summary	Title
Summary	Abstract		The abstract should allow the reader to determine the scope of the data collection. and accurately summarise its content.	Summary	Abstract
Summary	Publisher	Member Of	Please indicate if the organisation is an Alliance Member or a Hub.	*not part of Metadata Specs v1*	Publisher Default Attributes
Summary	Publisher	Access Rights	Organisation default access rights.	*not part of Metadata Specs v1*	Publisher Default Attributes
Summary	Publisher	Delivery Lead Time	Organisation default lead time.	*not part of Metadata Specs v1*	Publisher Default Attributes
Summary	Publisher	Data Use Limitation	Organisation default data use limitation.	*not part of Metadata Specs v1*	Publisher Default Attributes
Summary	Publisher	Data Use Requirements	Organisation default data use requirements.	*not part of Metadata Specs v1*	Publisher Default Attributes
Summary	Publisher	Access Service	Organisation default data access environment.	*not part of Metadata Specs v1*	Publisher Default Attributes
Summary	Publisher	Name	Organisation name.	Summary	Publisher
Summary	Publisher	Logo	Organisation logo.	*not part of Metadata Specs v1*	Publisher Default Attributes
Summary	Publisher	Description	Organisation description.	*not part of Metadata Specs v1*	Publisher Default Attributes
Summary	Publisher	Contact Point	Organisation default contact point.	*not part of Metadata Specs v1*	Publisher Default Attributes
Summary	Contact Point		Dataset contact point.	Summary	Contact Point
Summary	Keywords		Relevant and specific keywords that improve the SEO of the dataset.	Coverage	Keywords
Summary	Alternate Identifiers		Alternate dataset identifiers or local identifiers.	*not part of Metadata Specs v1*	
Summary	DOI		Unique Digital Object Identifier.	Attribution	DOI
Documentation	Description		A free-text description of the dataset.	Business	Description

Documentation	Associated Media		Additional context that could be useful for researchers wanting to understand more about the dataset.	*not part of Metadata Specs v1*	
Documentation	Is Part Of		Please complete only if the dataset is part of a group or family.	Summary	Group
Coverage	Spatial		The geographical area covered by the dataset.	Coverage	Geographic Coverage
Coverage	Typical Age Range		Please indicate the age range in whole years of participants in the dataset.	Coverage	Age Band
Coverage	Physical Sample Availability		If samples are available, please indicate the types of samples that are available.	Coverage	Physical Sample Availability
Coverage	Follow-Up		If known, what is the typical time span that a patient appears in the dataset (follow up period).	*not part of Metadata Specs v1*	
Coverage	Pathway		Please indicate if the dataset is representative of the patient pathway.	*not part of Metadata Specs v1*	
Provenance	Origin	Purpose	Pleases indicate the purpose(s) that the dataset was collected.	*not part of Metadata Specs v1*	
Provenance	Origin	Source	Pleases indicate the source of the data extraction.	*not part of Metadata Specs v1*	
Provenance	Origin	Collection Situation	Pleases indicate the setting(s) where data was collected.	*not part of Metadata Specs v1*	
Provenance	Temporal	Accrual Periodicity	Please indicate the frequency of publishing.	Coverage	Periodicity
Provenance	Temporal	Distribution Release Date	Date of the latest release of the dataset.	Business	Release Date
Provenance	Temporal	End Date	The end of the time period that the dataset provides coverage for.	Coverage	Dataset End Date
Provenance	Temporal	Time Lag	Please indicate the typical time-lag between an event and the data for that event appearing in the dataset.	*not part of Metadata Specs v1*	
Provenance	Temporal	Start Date	The start of the time period that the dataset provides coverage for.	Coverage	Dataset Start Date
Accessibility	Access	Access Rights	Description and guidance of the data access request process.	Summary	Access Rights
Accessibility	Access	Access Service	Brief description of the data access environment that is currently available to researchers.	*not part of Metadata Specs v1*	
Accessibility	Access	Delivery Lead Time	Indication of the typical processing data access request processing time.	Business	Access Request Duration
Accessibility	Access	Jurisdiction	Country/state under whose laws the data subjects' data is collected, processed and stored.	Coverage	Jurisdiction
Accessibility	Access	Data Controller	The person/entity who determines the purposes for which and the way any data, specifically personal data, are to be processed.	Business	Data Controller
Accessibility	Access	Data Processor	Any person/entity who processes the data on behalf of the data controller.	Business	Data Processor
Accessibility	Access	Access Request Cost	Indication of commercial model (in GBP) for processing each data access request by the data custodian .	Business	Access Request Cost
Accessibility	Usage	Data Use Limitation	Please provide an indication of consent permissions for the dataset.	*not part of Metadata Specs v1*	
Accessibility	Usage	Data Use Requirements	Please document restrictions and additional conditions for the use of the dataset.	*not part of Metadata Specs v1*	
Accessibility	Usage	Resource Creator	Text to be included as part of citations of credit for the dataset.	Attribution	Creator
Accessibility	Usage	Investigations		*not part of Metadata Specs v1*	

Accessibility	Usage	Is Referenced By	Please provide the keystone paper associated with the dataset.	Attribution	Citations
Accessibility	Formats and Standards	Vocabulary Encoding Scheme	Relevant terminologies / ontologies / controlled vocabularies that are being used by the dataset.	Structure	Controlled Vocabulary
Accessibility	Formats and Standards	Conforms To	Standardised data models that the dataset has been stored in or transformed to.	Structure	Conforms To
Accessibility	Formats and Standards	Language	Language in which dataset metadata and underlying data is made available.	Structure	Language
Accessibility	Formats and Standards	Format	File format of the dataset.	Structure	Format
Accessibility	Enrichment and Linkage	Qualified Relation	DOI of other datasets that have previously been linked to this dataset and their availability.	Business	Linked Datasets
Accessibility	Enrichment and Linkage	Derivation	Derived datasets and the type of derivation available or predefined extracts.	Business	Derived Datasets
Observations	Observed Node		The statistical populations for this observation.	*not part of Metadata Specs v1*	
Observations	Measured Value		The population size associated with the population type the dataset.	*not part of Metadata Specs v1*	
Observations	Observation Date		Date of the observation.	*not part of Metadata Specs v1*	
Observations	Disambiguating Description		Please provide a description that disambiguates the population type.	*not part of Metadata Specs v1*	
Observations	Measured Property		The measured property.	*not part of Metadata Specs v1*	
			refer to Data Use Limitations and Data Use Requirements	*Business*	License
			see Observations	*Coverage*	Statistical Population
			not part of Metadata Specs v2	*Structure*	File Size