Bioschemas Dataset - V0.2

[Summary](#_8x8b4ppa2gri)

[Contacts for the Specification](#_gmhv8ejoh3hl)

[Motivation](#_kd1f61ns4hgc)

[Use Cases](#_ngffjf4and1x)

[Specification](#_nm22w8uclka7)

[Mandatory Fields](#_72kz4mfy9l9c)

[Recommended Fields](#_46ngktw407wu)

[Examples](#_pqgsuh7qfom)

[Reactome dataset](#_wq7gwu755jv6)

[MetaboLights dataset](#_6h282x40iafm)

[ICPSR dataset](#_g422gux8frc3)

[Reference Material](#_ev79jdhlsr7c)

# Summary

This document describes a set of fields (properties, types and description) for discovering datasets, organized by their requirement level and driven by a set of use cases. The majority of these fields already exist in the [schema.org Dataset](http://schema.org/Dataset), others are pending. This V0.1 will be iteratively reviewed especially with the other Bioschemas groups.

# Contacts for the Specification

Philippe Rocca-Serra - [philippe.rocca-serra@oerc.ox.ac.uk](mailto:philippe.rocca-serra@oerc.ox.ac.uk)

Alejandra Gonzalez-Beltran - [alejandra.gonzalezbeltran@oerc.ox.ac.uk](mailto:alejandra.gonzalezbeltran@oerc.ox.ac.uk)

Susanna-Assunta Sansone - [susanna-assunta.sansone@oerc.ox.ac.uk](mailto:susanna-assunta.sansone@oerc.ox.ac.uk)

# Motivation

Most dataset repositories and registries do not provide structured data easily crawlable by search engines. Registries like DataMed, OMICsDI and BioSamples do automated ingestion of content mainly through APIs but not all the data repositories have a programmatic interface and the existing variety of programmatic interfaces are subject to changes which break integration workflows.

The purpose of this Bioschemas data group is to:

* Facilitate the ingestion of datasets metadata from data repositories (databases) into search engines and dataset registries like OMICsDI and DataMed via Bioschemas;
* Automate the linking of datasets metadata to samples in dataset registries like Biosamples, and identify cases where samples are missing or metadata is absent;
* Engage and help data providers to test and adopt the exposure of dataset metadata Bioschemas;
* Contribute to increase the number of indexed data repositories via Bioschemas;
* Make dataset registries compliant with Bioschemas.

# Use Cases

The overall aim is findability and not structured queries. Based on this, the key use cases that have driven the selection of the fields are the following:

* Search on a title of dataset
* Enable discovery by indexing on free text description
* Enable direct access, resolution of dataset
* Allow restriction to specific dimensions and variables specifically recorded in a dataset (e.g. get all climate datasets which have monitors CO2 concentration or datasets where metabolite concentration was recorded)
* Search according to the repository of the datasets
* Search according to creator, author
* Allow restriction to how a dataset has been generated (e.g. get all climate datasets if metabolite concentration was recorded if acquired using mass spectrometry)

# 

# Specification

The sections below outline the mandatory (minimum) and the recommended (optional) set of properties. These fields have been identified via a [crosswalk exercise](https://docs.google.com/spreadsheets/d/1XzrZxFIuG3TS9RU8vACoUjAvaADLmI_FrIk7O3BEkxY/edit?usp=sharing).

## Mandatory Fields

The mandatory fields are the minimum set (MUST) that is most relevant criteria to discover datasets, and within Bioschemas we have agreed to keep them to a maximum of six.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Mandatory Fields - MUST** | | | | | |
|  | ***Property*** | ***Expected Type*** | ***Cardinality*** | ***Description***  ***(source: schema.org Dataset descriptions in italics***  **with additional comments for Bioschemas*)*** | ***Presence in schema.org Dataset*** |
| 1 | [name](https://schema.org/name) | [Text](http://schema.org/Text) | *One* | *The name of the item.*  It is a descriptive name of the dataset | yes |
| 2 | [description](https://schema.org/description) | [Text](http://schema.org/Text) | *One* | *A description of the item.*  It is a short summary describing a dataset. | yes |
| 3 | [url](https://schema.org/url) | [URL](http://schema.org/URL) | *One* | *The URL of the item.*  It is the location of a page describing the dataset. | yes |
| 4 | [identifier](https://schema.org/identifier) | [PropertyValue](http://schema.org/PropertyValue) or [Text](http://schema.org/Text) or [URL](http://schema.org/URL) | *Many* | *The identifier property represents any kind of identifier for any kind of* [*Thing*](https://schema.org/Thing)*, such as ISBNs, GTIN codes, UUIDs etc. Schema.org provides dedicated properties for representing many of these, either as textual strings or as URL (URI) links.* It is recommended to be used for accession numbers or other identifiers apart from the URL, if they exist. Otherwise, make it equivalent to the URL. | yes |
| 5 | [keywords](https://schema.org/keywords) | [Text](http://schema.org/Text) | *Many* | *Keywords or tags used to describe this content. Multiple entries in a keywords list are typically delimited by commas.* These keywords provide a summary of the dataset. | yes |

## Recommended Fields

The recommended fields are optional (SHOULD), but providing them is encouraged because they enhance the discoverability.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Recommended Fields - SHOULD** | | | | |
|  | ***Property*** | ***Expected Type*** | ***Description***  ***(source: schema.org Dataset)*** | ***Presence in schema.org Dataset*** |
| 6 | [includedInDataCatalog](https://schema.org/includedInDataCatalog) | [DataCatalog](https://schema.org/DataCatalog) | *A data catalog which contains this dataset.* | yes |
| 7 | [creator](https://schema.org/creator) | [Text](http://schema.org/Text) | *The creator/author of this CreativeWork. This is the same as the Author property for CreativeWork.*  The name of the dataset creator (person or organization) | yes |
| 8 | [version](https://schema.org/version) | [Text](http://schema.org/Text), [Number](http://schema.org/Number) | The version number for this dataset | yes |
| 9 | [variableMeasured](http://pending.schema.org/variableMeasured) | [Text](http://schema.org/Text), [PropertyValue](http://schema.org/PropertyValue) | What does the dataset measure? (e.g., temperature, pressure) | [pending](http://pending.webschemas.org/variableMeasured) |
| 10 | [measurementTechnique](http://pending.schema.org/measurementTechnique) | [Text](http://schema.org/Text) | A technique or technology used for measuring the corresponding variable(s) (described using variablesMeasured) | [pending](http://pending.webschemas.org/measurementTechnique) |
| 11 | [citation](https://schema.org/citation) | [Text](http://schema.org/Text) | A citation for a publication that describes the dataset | yes |
| 12 | [license](https://schema.org/license) | [CreativeWork](https://schema.org/CreativeWork), [URL](http://schema.org/URL) | A license under which the dataset is distributed | yes |
| 13 | [distribution](http://schema.org/distribution) | [DataDownload](http://schema.org/DataDownload) | A downloadable form of this dataset, at a specific location, in a specific format | yes |

## Examples

### Reactome dataset

Status: in production

Available from: view-source:<http://reactome.org/content/detail/R-HSA-74160>

Use case: discovery

Documentation:<http://reactome.org/ContentService/#!/discover/eventDiscoveryUsingGET>

Implemented by Antonio Fabregat Mundo and Philippe Rocca-Serra during the Force11 Scholarly Communication Summer School hackathon, San Diego, August 2017.

|  |
| --- |
| <script type="application/ld+json">  **{**  **"name": "Gene expression (Transcription)",**  **"description": "Gene expression encompasses transcription and translation and the regulation of these processes. RNA Polymerase I Transcription produces the large preribosomal RNA transcript (45S pre-rRNA) that is processed to yield 18S rRNA, 28S rRNA, and 5.8S rRNA, accounting for about half the RNA in a cell. RNA Polymerase II transcription produces messenger RNAs (mRNA) as well as a subset of non-coding RNAs including many small nucleolar RNAs (snRNA) and microRNAs (miRNA). RNA Polymerase III Transcription produces transfer RNAs (tRNA), 5S RNA, 7SL RNA, and U6 snRNA. Transcription from mitochondrial promoters is performed by the mitochondrial RNA polymerase, POLRMT, to yield long transcripts from each DNA strand that are processed to yield 12S rRNA, 16S rRNA, tRNAs, and a few RNAs encoding components of the electron transport chain. Regulation of gene expression can be divided into epigenetic regulation, transcriptional regulation, and post-transcription regulation (comprising translational efficiency and RNA stability). Epigenetic regulation of gene expression is the result of heritable chemical modifications to DNA and DNA-binding proteins such as histones. Epigenetic changes result in altered chromatin complexes that influence transcription. Gene Silencing by RNA mostly occurs post-transcriptionally but can also affect transcription. Small RNAs originating from the genome (miRNAs) or from exogenous RNA (siRNAs) are processed and transferred to the RNA-induced silencing complex (RISC), which interacts with complementary RNA to cause cleavage, translational inhibition, or transcriptional inhibition.",**  **"url": "https://reactome.org/PathwayBrowser/#/R-HSA-74160",**  **"sameAs": null,**  **"version": "62",**  **"keywords": ["TopLevelPathway"],**  **"creator": [**  **{**  **"givenName": "Alberto R",**  **"familyName": "Kornblihtt",**  **"affiliation": [{**  **"name": "Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales",**  **"@type": "Organisation"**  **}],**  **"url": "http://europepmc.org/authors/null",**  **"@type": "Person"**  **},**  **{**  **"givenName": "Nicholas J",**  **"familyName": "Proudfoot",**  **"affiliation": [{**  **"name": "University of Oxford, Sir William Dunn School of Pathology",**  **"@type": "Organisation"**  **}],**  **"url": "http://europepmc.org/authors/null",**  **"@type": "Person"**  **},**  **{**  **"givenName": "Michael",**  **"familyName": "Caudy",**  **"affiliation": [{**  **"name": "Cold Spring Harbor Laboratory",**  **"@type": "Organisation"**  **}],**  **"url": "http://europepmc.org/authors/0000-0001-9459-4134",**  **"@type": "Person"**  **},**  **{**  **"givenName": "Peter",**  **"familyName": "D'Eustachio",**  **"affiliation": [{**  **"name": "NYU School of Medicine, Department of Biochemistry",**  **"@type": "Organisation"**  **}],**  **"url": "http://europepmc.org/authors/0000-0002-5494-626X",**  **"@type": "Person"**  **},**  **{**  **"givenName": "L",**  **"familyName": "Comai",**  **"affiliation": [],**  **"url": "http://europepmc.org/authors/0000-0003-3109-1841",**  **"@type": "Person"**  **},**  **{**  **"givenName": "Joan W",**  **"familyName": "Conaway",**  **"affiliation": [{**  **"name": "Stowers Institute for Medical Research",**  **"@type": "Organisation"**  **}],**  **"url": "http://europepmc.org/authors/null",**  **"@type": "Person"**  **},**  **{**  **"givenName": "Ron C",**  **"familyName": "Conaway",**  **"affiliation": [{**  **"name": "Stowers Institute for Medical Research",**  **"@type": "Organisation"**  **}],**  **"url": "http://europepmc.org/authors/null",**  **"@type": "Person"**  **},**  **{**  **"givenName": "Claes M",**  **"familyName": "Gustafsson",**  **"affiliation": [{**  **"name": "Karolinska Institute, Department of Laboratory Medicine",**  **"@type": "Organisation"**  **}],**  **"url": "http://europepmc.org/authors/null",**  **"@type": "Person"**  **},**  **{**  **"givenName": "Nouria",**  **"familyName": "Hernandez",**  **"affiliation": [{**  **"name": "Universite de Lausanne, Center for Integrative Genomics",**  **"@type": "Organisation"**  **}],**  **"url": "http://europepmc.org/authors/null",**  **"@type": "Person"**  **},**  **{**  **"givenName": "P",**  **"familyName": "Hu",**  **"affiliation": [{**  **"name": "University of California at Berkeley",**  **"@type": "Organisation"**  **}],**  **"url": "http://europepmc.org/authors/null",**  **"@type": "Person"**  **},**  **{**  **"givenName": "Nils-Goran",**  **"familyName": "Larsson",**  **"affiliation": [{**  **"name": "Karolinska Institute, Department of Laboratory Medicine",**  **"@type": "Organisation"**  **}],**  **"url": "http://europepmc.org/authors/null",**  **"@type": "Person"**  **},**  **{**  **"givenName": "Nicholas J",**  **"familyName": "Proudfoot",**  **"affiliation": [{**  **"name": "University of Oxford, Sir William Dunn School of Pathology",**  **"@type": "Organisation"**  **}],**  **"url": "http://europepmc.org/authors/null",**  **"@type": "Person"**  **},**  **{**  **"givenName": "Danny",**  **"familyName": "Reinberg",**  **"affiliation": [{**  **"name": "NYU School of Medicine, Department of Biochemistry",**  **"@type": "Organisation"**  **}],**  **"url": "http://europepmc.org/authors/null",**  **"@type": "Person"**  **},**  **{**  **"givenName": "Marc",**  **"familyName": "Timmers",**  **"affiliation": [{**  **"name": "University Medical Center Utrecht, Department of Physiological Chemistry",**  **"@type": "Organisation"**  **}],**  **"url": "http://europepmc.org/authors/null",**  **"@type": "Person"**  **}**  **],**  **"includedInDataCatalog": {**  **"url": "https://reactome.org",**  **"name": "Reactome",**  **"@type": "DataCatalog"**  **},**  **"distribution": [**  **{**  **"contentUrl": "https://reactome.org/ContentService/exporter/sbml/74160.xml",**  **"fileFormat": "SBML",**  **"@type": "DataDownload"**  **},**  **{**  **"contentUrl": "https://reactome.org/ReactomeRESTfulAPI/RESTfulWS/sbgnExporter/74160",**  **"fileFormat": "SBGN",**  **"@type": "DataDownload"**  **},**  **{**  **"contentUrl": "https://reactome.org/ReactomeRESTfulAPI/RESTfulWS/biopaxExporter/Level2/74160",**  **"fileFormat": "BIOPAX2",**  **"@type": "DataDownload"**  **},**  **{**  **"contentUrl": "https://reactome.org/ReactomeRESTfulAPI/RESTfulWS/biopaxExporter/Level3/74160",**  **"fileFormat": "BIOPAX3",**  **"@type": "DataDownload"**  **},**  **{**  **"contentUrl": "https://reactome.org/cgi-bin/pdfexporter?DB=gk\_current&ID=74160",**  **"fileFormat": "PDF",**  **"@type": "DataDownload"**  **},**  **{**  **"contentUrl": "https://reactome.org/cgi-bin/rtfexporter?DB=gk\_current&ID=74160",**  **"fileFormat": "DOCX",**  **"@type": "DataDownload"**  **},**  **{**  **"contentUrl": "https://reactome.org/cgi-bin/protegeexporter?DB=gk\_current&ID=74160",**  **"fileFormat": "OWL",**  **"@type": "DataDownload"**  **}**  **],**  **"citation": [],**  **"license": "https://creativecommons.org/licenses/by/4.0/",**  **"@context": "http://schema.org/",**  **"@type": "DataSet"**  **}**  </script> |

### MetaboLights dataset

Available at: <https://github.com/BioSchemas/_specifications/blob/master/Dataset/examples/mtbls234.json>

|  |
| --- |
| <script type="application/ld+json">  {  **"@context"**: **"http://schema.org/"**,  **"@type"**: **"Dataset"**,  **"identifier"**:**"MTBLS234"**,  **"url"**: **"http://www.ebi.ac.uk/metabolights/MTBLS234"**,  **"name"**: **"Automated Label-free Quantification of Metabolites from Liquid Chromatographyâ€“Mass Spectrometry Data (Plasma)Automated Label-free Quantification of Metabolites from Liquid Chromatographyâ€“Mass Spectrometry Data (Plasma)"**,  **"description"**: **"Liquid chromatography coupled to mass spectrometry (LC-MS) has become a standard technology in metabolomics. In particular, label-free quantification based on LC-MS is easily amenable to large-scale studies and thus well suited to clinical metabolomics. Large-scale studies, however, require automated processing of the large and complex LC-MS datasets. </p> We present a novel algorithm for the detection of mass traces and their aggregation into features (i.e. all signals caused by the same analyte species) that is computationally efficient and sensitive and that leads to reproducible quantification results. The algorithm is based on a sensitive detection of mass traces, which are then assembled into features based on mass-to-charge spacing, co-elution information, and a support vector machineâ€“based classifier able to identify potential metabolite isotope patterns. The algorithm is not limited to metabolites but is applicable to a wide range of small molecules (e.g. lipidomics, peptidomics), as well as to other separation technologies. </p> We assessed the algorithm's robustness with regard to varying noise levels on synthetic data and then validated the approach on experimental data investigating human plasma samples. We obtained excellent results in a fully automated data-processing pipeline with respect to both accuracy and reproducibility. Relative to state-of-the art algorithms, ours demonstrated increased precision and recall of the method. The algorithm is available as part of the open-source software package OpenMS and runs on all major operating systems. </p> See MTBLS235 for the simulated part of this study."**,  **"keywords"**: [  **"Mass spectrometry"**, **"Plasma"**,**"Spike-in"**,**"Control"**  ],  **"includedinDataCatalog"**: {  **"@type"**: **"DataCatalog"**,  **"name"**: **"EMBL-EBI Metabolights"**,  **"url"**: **"http://www.ebi.ac.uk/metabolights"**  },  **"creator"**:{  **"@type"**: **"Person"**,  **"familyName"**: **"Kohlbacher"**,  **"givenName"**: **"Oliver"**,  **"email"**: **"oliver.kohlbacher@uni-tuebingen.de"**,  **"affiliation"**: **"Wilhelm Schickard Institute for Computer Science, University of Tubingen"**  },  **"variablesMeasured"**: **"metabolite concentration"**,  **"measurementTechnique"**: **"mass spectrometry"**,  **"citation"**: {  **"@type"**: **"ScholarlyArticle"**,  **"headline"**: **"Automated Label-free Quantification of Metabolites from Liquid Chromatographyâ€“Mass Spectrometry Data"**,  **"identifier"**: [ **"http://doi.org/10.1074/mcp.M113.031278"**, **"https://www.ncbi.nlm.nih.gov/pubmed/24176773"**],  **"datePublished"**: **"2014-01"**  },  **"license"**: **"http://www.ebi.ac.uk/about/terms-of-use"**,  **"dateCreated"**: **"2015-09-23T"**,  **"datePublished"**: **"2015-10-06T"**,  **"dateModified"**: **"2016-12-15T"**,  **"fileFormat"**:**"ISA-Tab"**  }  </script> |

### ICPSR dataset

Available at: **http://doi.org/10.3886/ICPSR04652.v6**

|  |
| --- |
| {  **"@context"**: **"http://schema.org"**,  **"@type"**: **"Dataset"**,  **"identifier"**: **"http://doi.org/10.3886/ICPSR04652.v6"**,  **"url"**: **"http://doi.org/10.3886/ICPSR04652.v6"**,  **"name"**: **"Midlife in the United States (MIDUS 2), 2004-2006"**,  **"creator"**: [  {  **"@type"**: **"Person"**,  **"name"**: **"Ryff, Carol"**,  **"affiliation"**: **"University of Wisconsin-Madison"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Almeida, David M."**,  **"affiliation"**: **"Pennsylvania State University"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Ayanian, John"**,  **"affiliation"**: **"Harvard University"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Carr, Deborah S."**,  **"affiliation"**: **"University of Wisconsin-Madison"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Cleary, Paul D."**,  **"affiliation"**: **"Harvard University"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Coe, Christopher"**,  **"affiliation"**: **"University of Wisconsin-Madison"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Davidson, Richard"**,  **"affiliation"**: **"University of Wisconsin-Madison"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Krueger, Robert F."**,  **"affiliation"**: **"University of Minnesota"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Lachman, Marge E."**,  **"affiliation"**: **"Brandeis University"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Marks, Nadine F."**,  **"affiliation"**: **"University of Wisconsin-Madison"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Mroczek, Daniel K."**,  **"affiliation"**: **"Purdue University"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Seeman, Teresa"**,  **"affiliation"**: **"University of California-Los Angeles"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Seltzer, Marsha Mailick"**,  **"affiliation"**: **"University of Wisconsin-Madison"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Singer, Burton H."**,  **"affiliation"**: **"Princeton University"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Sloan, Richard P."**,  **"affiliation"**: **"Columbia University"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Tun, Patricia A."**,  **"affiliation"**: **"Brandeis University"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Weinstein, Maxine"**,  **"affiliation"**: **"Georgetown University"**  },  {  **"@type"**: **"Person"**,  **"name"**: **"Williams, David"**,  **"affiliation"**: **"University of Michigan"**  }  ],  **"version"**: **"6"**,  **"description"**: **"In 1995-1996, the MacArthur Midlife Research Network carried out a national survey of 7,108 Americans aged 25 to 74 (MIDLIFE IN THE UNITED STATES (MIDUS), 1995-1996 [ICPSR 2760]). The purpose of the study was to investigate the role of behavioral, psychological, and social factors in understanding age-related differences in physical and mental health. The study was innovative for its broad scientific scope, its diverse samples (which included twins and the siblings of main sample respondents), and its creative use of in-depth assessments in key areas (e.g., daily stress and cognitive functioning). A description of the study and findings from it are available at http://www.midus.wisc.edu. With support from the National Institute on Aging, a longitudinal follow-up of the original MIDUS samples: core sample (N = 3,487), metropolitan over-samples (N = 757), twins (N = 925 complete pairs), and siblings (N = 950), was conducted in 2004-2006. Guiding hypotheses for it, at the most general level, were that behavioral and psychosocial factors are consequential for physical and mental health. MIDUS II respondents were aged 35 to 86. Data collection largely repeated baseline assessments (e.g., phone interview and extensive self-administered questionnaire), with additional questions in selected areas (e.g., cognitive functioning, optimism and coping, stressful life events, and caregiving). To add refinements to MIDUS II, an African American sample (N = 592) was recruited from Milwaukee, Wisconsin, who participated in a personal interview and completed a questionnaire paralleling the above assessments. Also administered was a modified form of the mail questionnaire, via telephone, to respondents who did not complete a self-administered questionnaire."**,  **"includedInDataCatalog"**: {  **"@type"**: **"DataCatalog"**,  **"name"**: **"ICPSR General Archive"**,  **"url"**: **"http://www.icpsr.umich.edu/icpsrweb/ICPSR/"**  },  **"license"**: **"http://www.icpsr.umich.edu/icpsrweb/ICPSR/4652/terms"**,  **"keywords"**: [  **"adults"**,**"health status"**,**"life satisfaction"**,**"lifestyles"**,**"mental health"**,**"midlife"**,**"psychological wellbeing"**,**"relationships"**,**"siblings"**,**"social indicators"**,**"twins"**,**"work attitudes"**  ],  **"spatialCoverage"**: [  **"United States"**  ],  **"temporalCoverage"**: [  **"2004--2006"**  ],  **"funder"**: [  {  **"@type"**: **"Organization"**,  **"name"**: **"United States Department of Health and Human Services. National Institutes of Health. National Institute on Aging"**  }  ],  **"distribution"**: [  {  **"@type"**: **"DataDownload"**,  **"fileFormat"**: **"SAS"**,  **"contentURL"**: **"http://www.icpsr.umich.edu/cgi-bin/bob/terms2?study=4652&bundle=sas"**  },  {  **"@type"**: **"DataDownload"**,  **"fileFormat"**: **"SPSS"**,  **"contentURL"**: **"http://www.icpsr.umich.edu/cgi-bin/bob/terms2?study=4652&bundle=spss"**  },  {  **"@type"**: **"DataDownload"**,  **"fileFormat"**: **"Stata"**,  **"contentURL"**: **"http://www.icpsr.umich.edu/cgi-bin/bob/terms2?study=4652&bundle=stata"**  },  {  **"@type"**: **"DataDownload"**,  **"fileFormat"**: **"delimited"**,  **"contentURL"**: **"http://www.icpsr.umich.edu/cgi-bin/bob/terms2?study=4652&bundle=delimited"**  },  {  **"@type"**: **"DataDownload"**,  **"fileFormat"**: **"ASCII"**,  **"contentURL"**: **"http://www.icpsr.umich.edu/cgi-bin/bob/terms2?study=4652&bundle=ascii"**  }  ],  **"citation"**: **"Ryff, Carol, David M. Almeida, John Ayanian, Deborah S. Carr, Paul D. Cleary, Christopher Coe, Richard Davidson, Robert F. Krueger, Marge E. Lachman, Nadine F. Marks, Daniel K. Mroczek, Teresa Seeman, Marsha Mailick Seltzer, Burton H. Singer, Richard P. Sloan, Patricia A. Tun, Maxine Weinstein, and David Williams. Midlife in the United States (MIDUS 2), 2004-2006. ICPSR04652-v6. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2012-04-18. http://doi.org/10.3886/ICPSR04652.v6"**  } |

# Reference Material

Refer to the [crosswalk](https://docs.google.com/spreadsheets/d/1XzrZxFIuG3TS9RU8vACoUjAvaADLmI_FrIk7O3BEkxY/edit?usp=sharing) file where some of these are mapped to schema.org Dataset.

* [NIH BD2K DataMed DATS annotated with schema.org](https://biocaddie.org/group/working-group/working-group-3-descriptive-metadata-datasets)
* [OmicsDI model](http://biorxiv.org/content/early/2016/04/18/049205)
* [DataCite schemas](http://schema.datacite.org/)
* [A Data Citation Roadmap for Scholarly Data Repositories](http://biorxiv.org/content/early/2016/12/28/097196)
* [Automating the collection of “data repositories” metadata](https://docs.google.com/document/d/1N9-FTKL1GX2kmndIaSjbiJBxi0oykRYdJUjDkN2bFjE/edit)
* [Schema.org biological dataset via Bioschemas.org](https://docs.google.com/document/d/1N9-FTKL1GX2kmndIaSjbiJBxi0oykRYdJUjDkN2bFjE/edit)
* [Porting JATS into Schema.org](https://docs.google.com/document/d/1N9-FTKL1GX2kmndIaSjbiJBxi0oykRYdJUjDkN2bFjE/edit)
* [Google science dataset documentation](https://developers.google.com/search/docs/data-types/datasets)
* [Ontology-based Dataset Exploration](http://scim.brad.ac.uk/~dthakker/event/iesd2016/iesd16_submission_2.pdf)
* [The healthcare and life sciences community profile for dataset descriptions](https://peerj.com/articles/2331/)