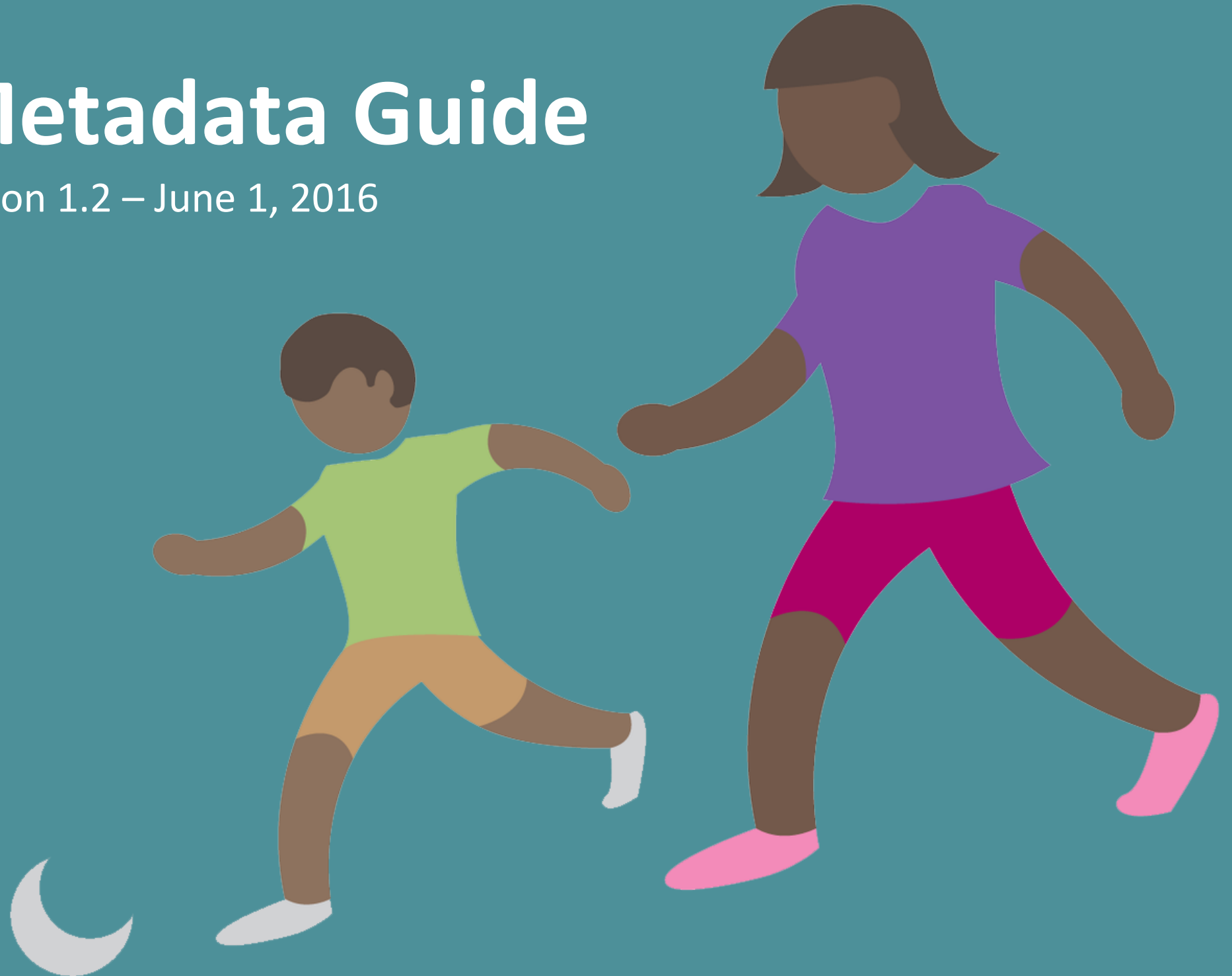




PolicyWise
for Children & Families

SAGE Metadata Guide

Version 1.2 – June 1, 2016



Revision History

Version	Date	Produced By	Change
1.0	May 16 2016	Lucie Richard	New guide, derived from prior, non DDI metadata documents proposed for SAGE
1.1	May 26 2016	Lucie Richard	Reformatted, added several sections and metadata field descriptors
1.2	June 1 st 2016	Lucie Richard	Add Field ID, match to metadata PDF products

Selected Glossary

The table below defines basic terms used throughout this document.

Term	Description
Affiliation	Academic institution, government agency or community organization to which an individual belongs or is associated at a given time
Coding Schema	Format or list of possible values that can be the content of a given metadata element
Controlled vocabulary	A standardized and organized arrangement of words and phrases used to represent content, enabling indexing and simple retrieval. Examples include Medical Subject Headings (MeSH), Library of Congress Subject Headings (LCSH), or Sears Subject Headings (SSH).
Dataset	A series of records with associated data elements describing them. Typically the elements are all related to a subject area or multiple subject areas, and might be restricted to a specific time frame.
Digital Object Identifier (DOI)	A serial code used to identify data assets uniquely
Element	A variable or characteristic collected about the record, usually constituting a column within a dataset
Obligation	The degree to which a metadata element is mandatory. In SAGE metadata, there are four levels: <ul style="list-style-type: none"> • Mandatory (M): the element must be populated in all instances • Mandatory if applicable (MA): the element must be populated if available and appropriate for the asset in question • Recommended (R): the element should be populated if at all possible • Optional (O): the element can be populated if available
Metadata	“Data about data”; Information which describes an element, dataset or study to provide an unfamiliar individual the necessary context to understand what the data is, its strengths and limitations, and how to appropriately use it without consulting the original researcher
Standard operating procedures	A document providing detailed procedures and instructions necessary to complete (in a standardized manner) data collection or other study-related activities
Study	One or more datasets that together enable a primary researcher to answer their research question(s). Also referred to as a Data Asset.

Table of Contents

Revision History	ii
Selected Glossary	iii
Table of Contents	iv
Using this Guide	1
Document-Level Metadata	2
Study-Level Metadata	3
Section A: Citation Information	3
Section B: Study Information	4
Section C: Methodology	7
Section D : Access and Terms of Use	8
Section E : Related Materials	9
Section F : SAGE Administrative Only	9
Dataset-Level Metadata	11
Section A : Dataset Description	11
Section B: Dataset Parameters	12
Section C: Access and Terms of Use	12
Element-Level Metadata	14
Section A: Element Description	14
Section B: Element Characteristics	14

Using this Guide

Metadata is a critical tool to increase the reach and legacy of data assets. All data assets within Secondary Analysis to Generate Evidence (SAGE) are equipped with metadata in order to facilitate data discovery, facilitate collaboration with other data repositories, and reduce the risk for inappropriate data interpretation or re-use. SAGE has elected to use metadata compatible with the Data Documentation Initiative (or DDI) standards (www.ddialliance.org), and is also compatible with Dublin Core Metadata Initiative (DCMI) and the Government of Alberta Core Content Standard (CORMAP).

Depositors may find this guide helpful while completing their data deposition with SAGE, as each field is listed with additional information. Users may also find this guide helpful while searching through SAGE data asset metadata, to better understand what a particular metadata field means. SAGE metadata fields are listed in this document with the following information:

Type of info	Description	Coding Scheme	Example
Name of field	The name of the metadata field	None	Study Producer Affiliation
Description	A description or definition of the metadata field	None	Affiliation of the study producer at the time of the study data acquisition
Type	The data type the metadata field represents	Text; Numeric; Date; Linked file	Text
Coding scheme	Lists any specific coding scheme mandated by the metadata element	None	0 = Complete; 1 = Unverified; 2 = Incomplete
Obligation	The degree to which the metadata field is mandatory	Mandatory (M) Mandatory if applicable (MA) Recommended (R) Optional (O) ¹	M
Repeatable	Whether the metadata field can be repeated (eg. When there are multiples of the same elements – as with data producers, funders, etc.)	Yes; No	Yes
Created By	The individual or organization responsible for populating the metadata field	Depositor; SAGE staff	Depositor
Intended Use	The type of intended use of the metadata field	Descriptive (D); Administrative (A); Use (U) ²	Descriptive
Example	Provides an example of the metadata element	None	University of Calgary

The following sections describe the metadata fields for four levels of SAGE metadata: document level; study level; dataset level; and element level.

¹ Mandatory (M) fields are considered the critical minimal set for discovery;

Mandatory if Applicable (MA) fields are considered critical for data assets where they are applicable only (ie. If it exists for the data asset, it must be disclosed in metadata);

Recommended (R) fields are considered highly desirable where applicable but not critical; and

Optional (O) fields are considered desirable where applicable, but optional

² Descriptive (D) fields are mainly intended for data discovery – they help describe the resource and how it is to be used. “Classical metadata”;

Administrative (A) fields provide SAGE with internal information about the resource, such as agreement renewal dates, storage locations etc. They are generally not available to external users.

Use (U) fields describe how the resource has been used within the repository, such as number of people who have accessed the data asset.

Document-Level Metadata

Document-level metadata are meant to describe the documentation for the study (as separate from the data itself): ie. It is metadata about the documentation available. There are 18 document-level elements.

Field ID	Name of Field	Description	Type	Coding Scheme	Obligation	Repeatable	Created By	Intended Use	Example
00001	Title	Title of the metadata document: should include title of study/dataset it describes	Text	None	Mandatory	No	Depositor	Descriptive	Documentation for the All Our Babies study data (2008-2013) deposited with Secondary Analysis to Generate Evidence
00002	Identifier	A unique identifier for the metadata document	Text	S[XX]_[abtitle]_metadata_v[XX]	Mandatory	No	SAGE Staff	Descriptive	S01_AOB_metadata_v01
00003	Author	Author with accountability for the metadata content	Text	Lastname, Firstname	Mandatory	Yes	Depositor	Descriptive	Tough, Suzanne
00004	Author – Affiliation	Affiliation of the author at the time of the metadata production	Text	None	Mandatory if Applicable	Yes	Depositor	Descriptive	University of Calgary
00005	Producer	Agency or person who prepared the metadata document in its current form	Text	None	Mandatory	No	Depositor	Descriptive	Secondary Analysis to Generate Evidence
00006	Distributor	Known distributor of this data asset’s metadata	Text	None	Mandatory	Yes	Depositor	Descriptive	Secondary Analysis to Generate Evidence
00007	Distributor - Affiliation	Known distributor affiliation, if applicable	Text	None	Mandatory if Applicable	Yes	Depositor	Descriptive	Alberta Centre for Child, Family and Community Research
00008	Distributor URI	Known distributor Uniform Resource Identifier (URI)	Text	None	Mandatory if Applicable	Yes	Depositor	Descriptive	http://www.Research4children.com
00009	Contact	Person or organization who can be reached for questions regarding the data asset metadata	Text	None (if person: lastname, first name)	Mandatory	Yes	Depositor	Descriptive	Tough, Suzanne
00010	Contact - Affiliation	Affiliation of the contact person or organization	Text	None	Mandatory if Applicable	Yes	Depositor	Descriptive	University of Calgary
00011	Contact - Email	Email of the contact person or organization	Text	[XXXXX]@[XXXXX].[XX]	Mandatory	Yes	Depositor	Descriptive	stough@ucalgary.ca
00012	Contact - URI	Uniform Resource Identifier (URI) of contact person or organization. This URI should preferably relate to the data asset in some way	Text	None	Mandatory if Applicable	Yes	Depositor	Descriptive	https://www.ucalgary.ca/stough/allourbabies
00013	Document version	Version of this data asset’s metadata	Text	v. [X].[X]	Mandatory	No	Depositor	Descriptive	v. 1.0
00014	Version date	Date this version of the metadata was completed	Date	YYYY/MM/DD	Mandatory	No	Depositor	Descriptive	2016/04/25
00015	Version notes	Any additional information regarding the version of the metadata, in particular what makes a version different from its predecessor	Text	None	Recommended	No	Depositor	Descriptive	
00016	Metadata bibliographic citation	Complete bibliographic citation for the data asset metadata	Text	[Producer]; [version year]; “[Title]”; [DOI]; [Distributor]; [Version]	Mandatory	No	SAGE Staff	Descriptive	Secondary Analysis to Generate Evidence; 2016, "Documentation for the All Our Babies study data (2008-2013) deposited with Secondary Analysis to Generate Evidence", < http://doi:10.7939/DVN/10793 >, Secondary Analysis to Generate Evidence, v1
00017	Location	Internal storage location of the data asset metadata	Text	None	Mandatory	No	SAGE Staff	Administrative	http://matrix.otcs/llisapi.dll/open/578776
00018	Document Status	Indicates the degree of completeness of the data asset metadata	Text	0=Incomplete; 1=Unverified; 2=Complete	Mandatory	No	SAGE Staff	Descriptive	1

Study-Level Metadata

Study-level metadata are meant to describe the overall data asset (or overall study), which may include multiple instruments or datasets. It is by far the largest part of SAGE metadata: There are 82 elements in 6 sections.

Field ID	Name of Field	Description	Type	Coding Scheme	Obligation	Repeatable	Created By	Intended Use	Example
<i>SECTION A : CITATION INFORMATION</i>									
00019	Title	Full, official title of the data asset. May be the same as listed in the funding grant, if applicable	Text	None	Mandatory	No	Depositor	Descriptive	All Our Babies Study
00020	Subtitle	A secondary title used to amplify or state certain limitations upon the official title	Text	None	Recommended	No	Depositor	Descriptive	Developing a unique research legacy to inform policy, practice, and health service planning to improve outcomes for Alberta's children
00021	Alternative title	Alternative title, eg. The title by which a data collection is commonly referred to, or an abbreviation	Text	([xxx])	Recommended	No	Depositor	Descriptive	AOB
00022	Identifier	A unique identifier for the data asset	Text	S[XX]_[abbtile]_v[XX]	Mandatory	No	SAGE Staff	Descriptive	S01_AOB_v1
00023	DOI	Data asset digital object identifier (DOI)	DOI string	doi:[xx.xxxx]/[xxx]/[xxxxx]	Mandatory	No	SAGE Staff	Descriptive	doi:10.7939/DVN/10793
00024	Author	Person, corporate body or organization responsible for the data asset's substantive or intellectual content	Text	If name: lastname, first name	Mandatory	Yes	Depositor	Descriptive	Tough, Suzanne
00025	Author - affiliation	Affiliation of the author	Text	None	Mandatory if Applicable	Yes	Depositor	Descriptive	University of Calgary
00026	Producer	Person or organization with financial or administrative responsibility for the data asset (in the case of research data, typically the investigator leading a grant and/or ethics application)	Text	If name: lastname, first name	Mandatory	Yes	Depositor	Descriptive	Tough, Suzanne
00027	Producer – role	Role of producer in the study	Text	None	Mandatory if Applicable	Yes	Depositor	Descriptive	Principal investigator
00028	Producer - Affiliation	Affiliation of producer at the time of the asset creation	Text	None	Mandatory if Applicable	Yes	Depositor	Descriptive	University of Calgary
00029	Distributor	Organization that disseminates this data asset outside the original study team	Text	None	Mandatory	Yes	Depositor	Descriptive	Secondary Analysis to Generate Evidence
00030	Distributor – Affiliation	Affiliation of data asset distributor	Text	None	Recommended	Yes	Depositor	Descriptive	Alberta Centre for Child, Family and Community Research
00031	Distributor URI	Uniform Resource Identifier (URI) of data asset distributor	Text	None	Recommended	Yes	Depositor	Descriptive	http://www.Research4children.com
00032	Contact	Person or organization who can be reached for questions regarding the data asset (as distinct from the data asset metadata)	Text	If name: lastname, firstname	Mandatory	Yes	Depositor	Descriptive	Wu, Muci
00033	Contact - Email	Email of study contact at the time of metadata creation	Text	[XXXXX]@[XXXXX].[XX]	Mandatory	Yes	Depositor	Descriptive	muci.wu@ucalgary.ca
00034	Contact –Affiliation	Affiliation of study contact at time of metadata creation	Text	None	Recommended	Yes	Depositor	Descriptive	University of Calgary

Name of Field	Description	Type	Coding Scheme	Obligation	Repeatable	Created By	Intended Use	Example	
00035	Contact – URI	Uniform Resource Identifier (URI) of study contact at time of metadata creation	Text	None	Recommended	Yes	Depositor	Descriptive	http://allourbabiesstudy.com/staff/
00036	Funding source	Organization that funded the creation of the data asset	Text	None	Mandatory if Applicable	Yes	Depositor	Descriptive	Alberta Innovates – Health Solutions;
00037	Funding grant	Name and grant number that supported the creation of the data asset	Text	None	Optional	Yes	Depositor	Descriptive	Alberta Heritage Foundation for Medical Research Preterm Birth and Healthy Outcomes #200700595
00038	Funding intent	What aspect of the creation of the data asset the funding source supported	Text	None	Optional	Yes	Depositor	Descriptive	Equipment
00039	Data deposit agreement	Data deposit agreement between the data asset owner and SAGE	LINKED FILE	None	Mandatory	No	SAGE Staff	Administrative	http://matrix.otcs/llisapi.dll/open/378511
00040	Data deposit agreement - date	Date when the deposit agreement was signed (date of last signature)	Date	YYYY/MM/DD	Mandatory	No	SAGE Staff	Administrative	2015/09/03
00041	Location	Internal storage location of the deposited data asset	Text	None	Mandatory	No	SAGE Staff	Administrative	G:/External/USR/MWU/
00042	Storage space	Storage space required for all data (unzipped, in MB)	Number	None	Mandatory	No	SAGE Staff	Descriptive	22.6
00043	Bibliographic citation	How this data asset should be cited in secondary research	Text	[Author(s)]; [year created]; “[Title]”; [DOI]; [Distributor]; [Version]	Mandatory	No	SAGE Staff	Descriptive	Tough, Suzanne; Graham, Susan; 2013, "All Our Babies", http://doi:10.7939/DVN/10793 , Secondary Analysis to Generate Evidence,
<i>SECTION B : STUDY DESCRIPTION</i>									
00044	Number of datasets	Number of datasets in data asset	Number	Integer	Mandatory	No	Depositor	Descriptive	13
00045	Holdings	Brief description of the contents of the data asset. Typically includes, at a minimum, the title of the dataset(s) included, in chronological order	Text	None	Recommended	No	Depositor	Descriptive	This deposit of AOB data includes the following datasets: 1. Q1M (Questionnaire 1, Maternal) - First maternal questionnaire, given at <25 weeks gestation; 2. Q2M (Questionnaire 2, Maternal) - Second maternal questionnaire, given at 34-36 weeks gestation; 3. Q3M (Questionnaire 3, Maternal) - Third maternal questionnaire, given at 4 months post-partum; 4. Q4M (Questionnaire 4, Maternal) - Fourth maternal questionnaire, given at 12 months post-partum; ...
00046	Keywords	Keywords describing the salient aspects of the study content. A controlled vocabulary may be employed	Text	None	Mandatory	Yes	Depositor	Descriptive	Prenatal influences

	Name of Field	Description	Type	Coding Scheme	Obligation	Repeatable	Created By	Intended Use	Example
00047	Keywords – Vocabulary	Indicate controlled vocabulary used, if used	Text	None	Optional	Yes	Depositor	Descriptive	Library of Congress Subject Headings
00048	Study Summary	A short summary describing the purpose, nature and scope of the data asset, written for a broader scientific audience. May include special characteristics of its contents, major subject areas covered, and/or what the investigators aimed to answer when they created the asset.	Text	None	Mandatory	No	Depositor	Descriptive	The All Our Babies (AOB) Study is a prospective pregnancy cohort study designed to enable investigation of the relationship between maternal and family well-being and outcomes for children. AOB participants have provided information about demographics, mental health, lifestyle, pregnancy, parenting, child development, family life and service use to identify early markers of later outcomes. Subject areas covered include: a) Maternal Prenatal Health (e.g., prenatal care, maternal gestation); b) Pregnancy History ...
00049	Collection Mode	The method(s) used to collect the data in this data asset (list all data methods used up to time of data deposit <i>that are included in the data deposit</i>)	Text	None	Mandatory	No	Depositor	Descriptive	Questionnaire
00050	Study design	Overall study design	Text	0=Cross-sectional 1=Qualitative 2=Cohort 3=Case-control 4=Quasi-experimental 5=Mixed/other	Mandatory	No	Depositor	Descriptive	2
00051	Study design – Other	If [Study design]=5, please describe	Text	None	Mandatory if Applicable	No	Depositor	Descriptive	Randomized controlled trial
00052	Time dimension	Time method or time dimension of data collection	Text	0=Longitudinal (same individuals over time); 1=Time series (different individuals over time); 2=Cross-sectional (individuals at a single point in time);	Mandatory	No	Depositor	Descriptive	0
00053	Time period – start	Start of time period to which the data asset refers (time covered by the data asset information, NOT necessarily the same as dates of collection)	Date	YYYY/MM/DD OR YYYY/MM	Recommended	No	Depositor	Descriptive	2007/11/08
00054	Time period – end	End of time period to which the data asset refers (time covered by the data asset information, NOT necessarily the same as dates of collection)	Date	YYYY/MM/DD OR YYYY/MM	Recommended	No	Depositor	Descriptive	2013/12/30
00055	Collection start date	Date when earliest data collection began	Date	YYYY/MM/DD OR YYYY/MM	Mandatory	No	Depositor	Descriptive	2008/05/12
00056	Collection end date	Date when latest data collection ended	Date	YYYY/MM/DD OR YYYY/MM	Mandatory	No	Depositor	Descriptive	2011/05/19
00057	Countries	List of countries in which the study was undertaken	Text	[Country1]; [Country2]	Mandatory	No	Depositor	Descriptive	Canada

Name of Field	Description	Type	Coding Scheme	Obligation	Repeatable	Created By	Intended Use	Example	
00058	Geographic coverage	List of cities or administrative regions in which the study was undertaken. Please include the province/territory/state.	Text	None	Mandatory	No	Depositor	Descriptive	Calgary and surrounding area, Alberta
00059	Study Language	Language in which the study was conducted. English is assumed if left empty. Enter all languages where study used more than one language. Enter N/A where study used no language.	Text	None	Recommended	No	Depositor	Descriptive	English
00060	Unit of analysis	Data asset contains information collection at which level(s) of aggregation (eg. Individual, family, neighbourhood)	Text	None	Recommended	No	Depositor	Descriptive	Maternal infant/child dyad
00061	Study universe	Description of the type of persons who are the object of research and to which any analytic results refer. Description may include any number of factors, such as age, nationality, residence, gender, race, income criminal convictions, etc.	Text	None	Mandatory	No	Depositor	Descriptive	Pregnant women over the age of 18 who are able to communicate in written and spoken English and living in Canada
00062	Eligibility Criteria	Description of the criteria participants must meet to be included in the study	Text	None	Mandatory	No	Depositor	Descriptive	Women who are 1) over the age of 18; b) are fewer than 25 weeks pregnant; c) are able to communicate in written and spoken English
00063	Eligibility – age min	List minimum age to be eligible for inclusion in the study	Number	Integer	Mandatory if Applicable	No	Depositor	Descriptive	18
00064	Eligibility – age max	List maximum age to be eligible for inclusion in the study, if available	Number	Integer	Mandatory if Applicable	No	Depositor	Descriptive	105
00065	Eligibility – gender	List gender(s) eligible for inclusion in the study	Text	None	Mandatory if Applicable	No	Depositor	Descriptive	Female
00066	Stratification	Was there any sample or data stratification (eg. By region, treatment type, hospital)	Text	0=No; 1=Yes; 2=N/A	Recommended	No	Depositor	Descriptive	0
00067	Stratification - type	Description of stratification if present. Include information about any changes to stratification over the study period	Text	None	Recommended	No	Depositor	Descriptive	Stratified by geographic region (Edmonton and area; Calgary and area; Other)
00068	Reasons for dropout	Describe known causes of dropout from study, other than loss of interest or lost to follow up	Text	None	Recommended	No	Depositor	Descriptive	Discontinuation/loss of pregnancy
00069	Incentives for participation	Incentives for participation in study (may list multiple)	Text	0 = None 1 = Gift Card 2 = Cash 3 = Entry into draw or lottery 4 = Other	Mandatory	No	Depositor	Descriptive	1; 3
00070	Incentives– Other	If incentives = 4, please describe	Text	None	Mandatory if Applicable	No	Depositor	Descriptive	Free dinner during focus group session
00071	Linkage key	List of any keys used to link data across data asset	Text	None	Mandatory if Applicable	No	Depositor	Descriptive	StudyID
00072	Interoperability	List any standardization used to enhance comparability between data assets (eg. Standardized instruments used)	Text	None	Recommended	No	Depositor	Descriptive	Standardized instruments used include: Perceived Stress Scale; ...

Name of Field	Description	Type	Coding Scheme	Obligation	Repeatable	Created By	Intended Use	Example		
<i>SECTION C : METHODOLOGY</i>										
00073	Standard operating procedures – link	Provide any standard operating procedures used for to create this data asset	Linked File	None		Optional	No	Depositor	Descriptive	.
00074	Recruitment procedure	Description of the method(s) by which the study sample was acquired	Text	None		Recommended	No	Depositor	Descriptive	Recruitment occurred in two ways : 1. Posters and word-of-mouth advertisement at health care offices; the regional health services website 2. Women receiving prenatal viral serology tests were asked for permission to be contacted by research staff to discuss participation and determine eligibility
00075	Sampling procedure	Description of the method(s) by which the cohort was sampled from the population. If no method used please state “No sampling procedure used”	Text	None		Mandatory	No	Depositor	Descriptive	No sampling procedure used
00076	Representativeness of sample	Describe correspondence (and any discrepancies) between sampled units and available statistics for the universe the sample is meant to represent	Text	None		Recommended	No	Depositor	Descriptive	Compared to Maternity Experiences Survey (Canadian cross-sectional survey), AOB mothers were older (>35 years), reported a higher household income, were more likely to have received a first ultrasound before 18 weeks gestational age and to attend prenatal or childbirth education classes. Otherwise, AOB and MES participant demographic and pregnancy characteristics were similar. Additional information available in McDonald, S. W., Lyon, A. W., Benzies, K. M., McNeil, D. A., Lye, S. J., Dolan, S. M., ... & Tough, S. C. (2013). <i>The All Our Babies pregnancy cohort: design, methods, and</i>
00077	Factors affecting data collection	Describe any noteworthy aspects of data collection, such as cooperativeness of respondents, duration of interviews, number of call-backs, etc.	Text	None		Optional	No	Depositor	Descriptive	Duration of interview averaged 27 minutes
00078	Actions to minimize data loss	Describe any actions taken to minimize loss to follow-up (eg. Follow up visits, supervisory checks, historical matching, estimation)	Text	None		Optional	No	Depositor	Descriptive	.
00079	Weighting	Describe variable, method or formula used to apply weights to produce accurate statistical results. If no weighting used, state “No weighting used”	Text	None		Recommended	No	Depositor	Descriptive	No weighting used
00080	Pre-deposition cleaning	Description of any methods used to clean or preprocess the study datasets prior to data deposition	Text	None		Recommended	No	Depositor	Descriptive	.
00081	Sample size – recruited	Overall sample size recruited for data asset (actual, not targeted)	Number	Integer		Recommended	No	Depositor	Descriptive	4011

Name of Field	Description	Type	Coding Scheme	Obligation	Repeatable	Created By	Intended Use	Example
00082	Sample size – collected	Number	Integer	Mandatory	No	Depositor	Descriptive	3390
00083	Response rate	Number	Percentage	Optional	No	Depositor	Descriptive	26.1%
00084	Study Status	Text	0=No; 1=Yes; 2=N/A	Mandatory	No	Depositor	Descriptive	1
<i>SECTION D : ACCESS AND TERMS OF USE</i>								
00085	Original archive	Text	None	Recommended	No	Depositor	Descriptive	Population Research Lab (Department of Sociology, University of Alberta)
00086	Completeness of deposit	Text	None	Mandatory if Applicable	No	Depositor	Descriptive	<ol style="list-style-type: none"> Blood samples (optional) taken at 18-22 weeks gestation, and 28-32 weeks gestation Cord blood, collected at birth Survey at 5 years postpartum (data not ready) Alberta Health Services electronic health medical record
00087	Availability	Text	0= No; 1= Yes	Mandatory	No	Depositor	Descriptive	1
00088	Access conditions	Text	None	Mandatory	No	SAGE Staff	Descriptive	No special access procedures or conditions
00089	Access Restrictions	Text	None	Mandatory	No	SAGE Staff	Descriptive	36 month data currently under embargo
00090	Recognition requirement	Text	None	Mandatory if Applicable	No	SAGE Staff	Descriptive	<p>Data must be cited in any publication or report using data asset bibliographic citation</p> <p>Co-authorship required where scientifically appropriate (as determined by publication authorship guidelines)</p>
00091	Data security level	Text	1 = Level 1 – access only through physical enclave; 2 = Level 2 – access using Virtual Research Environment; 3 = Level 3 – access via secure FTP 4 = Level 4 – Open Access	Mandatory	No	SAGE Staff	Descriptive	2

	Name of Field	Description	Type	Coding Scheme	Obligation	Repeatable	Created By	Intended Use	Example
00092	Vetting requirements	Any vetting requirements to export results from SAGE for reporting and publication that differs from the standard SAGE vetting requirements checklist. If none applicable, state "No special vetting requirements"	Text	None	Mandatory	No	SAGE Staff	Descriptive	No special vetting requirements
<i>SECTION E : RELATED MATERIALS</i>									
00093	Related material	Citations (in APA format) of published papers, reports and other references key to understanding the design or methodology of the study.	Text	APA format	Mandatory if Applicable	No	Depositor	Descriptive	<p>Gracie, S. K., Lyon, A. W., Kehler, H. L., Pennell, C. E., Dolan, S. M., McNeil, D. A., ... & Hegadoren, K. M. (2010). All Our Babies Cohort Study: recruitment of a cohort to predict women at risk of preterm birth through the examination of gene expression profiles and the environment. BMC pregnancy and childbirth, 10(1), 87.;</p> <p>McDonald, S. W., Lyon, A. W., Benzies, K. M., McNeil, D. A., Lye, S. J., Dolan, S. M., ... & Tough, S. C. (2013). The All Our Babies pregnancy cohort: design, methods, and participant characteristics. BMC pregnancy and childbirth, 13(1), 1. ;</p> <p>...</p>
00094	Results	Citations (in APA format) of published papers, reports and other references that describe research questions that were answered using the data asset.	Text	APA format	Mandatory	No	Depositor	Descriptive	<p>Nagulesapillai, Tharsiya, Sheila W. McDonald, Tanis R. Fenton, H. F. Mercader, and Suzanne C. Tough. "Breastfeeding difficulties and exclusivity among late preterm and term infants: results from the all our babies study." Canadian Journal of Public Health 104, no. 4 (2013): e351-e356.;</p> <p>Bat-Erdene, U., Metcalfe, A., McDonald, S. W., & Tough, S. C. (2013). Validation of Canadian mothers' recall of events in labour and delivery with electronic health records. BMC pregnancy and childbirth, 13(1), 1.;</p> <p>...</p>
<i>SECTION F : SAGE ADMINISTRATIVE ONLY</i>									
00095	Data preservation	Description of any information pertaining to long-term preservation of the data. Should include any obligation to destroy or return and conditions thereof.	Text	None	Mandatory if Applicable	No	SAGE Staff	Administrative	.

	Name of Field	Description	Type	Coding Scheme	Obligation	Repeatable	Created By	Intended Use	Example
00096	Special clauses	Special parameters noted in Data Deposit Agreement that impact intellectual property, access, storage, data processing, linkage, or metadata in a way not otherwise documented in the data asset metadata	Text	None	Mandatory if Applicable	No	SAGE Staff	Administrative	
00097	Ethics approvals for deposit	Ethics approval(s) for deposit to SAGE	Linked File	None	Mandatory	No	SAGE Staff	Administrative	http://matrix.otcs/llisapi.dll/open/55937 ; http://matrix.otcs/llisapi.dll/open/55936
00098	Data access agreements	Links to any data access agreements that have accessed this data.	Linked File	None	Mandatory if Applicable	No	SAGE Staff	Use	http://matrix.otcs/llisapi.dll/open/55572
00099	Contingency for breach	Any special procedures above and beyond standard SAGE procedures in the event of breach	Text	None	Mandatory if applicable	No	SAGE Staff	Administrative	
00100	Marketing messages	Key marketing and communications messages for this study, such as advantages, unique attributes, etc.	Text	None	Mandatory if Applicable	No	Depositor	Administrative	High response rate; Birth cohort; Comprehensive investigation of topics relating to early life child development

Dataset-Level Metadata

Dataset level metadata are meant to capture information for a particular dataset within a study. There are 32 elements in 3 sections.

Field ID	Name of Field	Description	Type	Coding Scheme	Obligation	Repeatable	Created By	Intended Use	Example
<i>SECTION A : DATASET DESCRIPTION</i>									
00101	Title	Descriptive name of dataset	Text	None	Mandatory	No	Depositor	Descriptive	Questionnaire 6 – Maternal
00102	Alternative title	Alternative title of the dataset, the title by which a data collection is commonly referred to, or an abbreviation	Text	([xxx])	Mandatory if Applicable	No	Depositor	Descriptive	Q6M
00103	Identifier	A unique identifier for the dataset	Text	S[XX]_[abbtile]_ [Alttile]	Mandatory	No	SAGE Staff	Descriptive	S01_AOB_Q6M
00104	Version	What is the version of this dataset	Text	V [x].[x]	Mandatory	No	Depositor	Descriptive	v 1.1
00105	Version Date	Date of this version of the dataset	Date	YYYY/MM/DD OR YYYY/MM	Mandatory	No	Depositor	Descriptive	2015/09/03
00106	Version Notes	Describe in what way this dataset version differs from predecessors (why it was altered)	Text	None	Recommended	No	Depositor	Descriptive	Data errors found in v 1.0 were corrected
00107	Content	A short summary about the dataset. In studies with multiple datasets, indicate this dataset’s relationship to the other datasets.	Text		Mandatory	No	Depositor	Descriptive	The sixth maternal questionnaire, given at 36 months, designed to follow up with participants on previously collected data examining determinants of maternal, infant, and child outcomes. Is roughly coincident with the 3 rd CDI questionnaire (Q6CDI) and 3 rd ASQ (Q6ASQ); Follows the 5 th maternal questionnaire
00108	Producer	Principal investigator of original study or investigator with primary accountability for the dataset, IF DIFFERENT FROM STUDY LEVEL	Text	Lastname, Firstname	Mandatory if Applicable	Yes	Depositor	Descriptive	.
00109	Producer - role	Role of producer	Text	None	Mandatory if Applicable	Yes	Depositor	Descriptive	.
00110	Producer - Affiliation	Affiliation of producer at the time of the dataset collection	Text	None	Recommended	Yes	Depositor	Descriptive	.
00111	Sample size	Sample size of dataset	Number	Integer	Mandatory	No	Depositor	Descriptive	1994
00112	Number of records	Number of records in dataset, if different than sample size	Number	Integer	Mandatory if Applicable	No	Depositor	Descriptive	2410
00113	Number of data elements	Number of columns/variables/data elements	Number	Integer	Mandatory	No	Depositor	Descriptive	118
00114	Elements excluded	List of elements excluded from dataset in this deposit (eg. due to rare events, privacy concerns).	Text	None	Mandatory if Applicable	No	Depositor	Descriptive	Participant date of birth
00115	Data transformations	Describe any dataset transformations and reasons for transformations (eg. Privacy). Also provide methodology used to transform data, if applicable.	Text	None	Recommended	No	Depositor	Descriptive	Participant age replaces participant date of birth (excluded from data asset).
00116	File format	File format of the dataset when deposited	Text	None	Mandatory	No	SAGE Staff	Administrative	.sav (SPSS)

	Name of Field	Description	Type	Coding Scheme	Obligation	Repeatable	Created By	Intended Use	Example
00117	Location	Internal storage location of the deposited dataset	Text	None	Mandatory	No	SAGE Staff	Administrative	G:/External/USR/MWU/
<i>SECTION B: DATASET PARAMETERS</i>									
00118	Time Period – start	Start of time period referred to by the information in the dataset (Not necessarily the same as the dataset collection start date).	Date	YYYY/MM/DD OR YYYY/MM	Recommended	No	Depositor	Descriptive	2012/03/08
00119	Time Period - end	End of time period referred to by the information in the dataset (Not necessarily the same as the dataset collection end date).	Date	YYYY/MM/DD OR YYYY/MM	Recommended	No	Depositor	Descriptive	2014/12/02
00120	Collection - start date	Start date of the dataset collection	Date	YYYY/MM/DD OR YYYY/MM	Mandatory	No	Depositor	Descriptive	2012/03/08
00121	Collection - end date	End date of the dataset collection	Date	YYYY/MM/DD OR YYYY/MM	Mandatory	No	Depositor	Descriptive	2014/12/02
00122	Collection status	Whether dataset collection is complete at time of deposit	Text	0=No; 1=Yes;	Mandatory	No	Depositor	Descriptive	1
00123	Geographic coverage	List of cities or administrative regions in which the dataset collection was undertaken. Include the province, territory or state where the city or administrative region exists.	Text	None	Mandatory	No	Depositor	Descriptive	Calgary and surrounding area, Alberta
00124	Unit of analysis	Dataset contains information collected at which level(s) of aggregation (eg. Individual, family, neighbourhood),	Text	None	Mandatory	No	Depositor	Descriptive	Maternal infant/child dyad
00125	Eligibility	Eligibility requirement for data to be collected for that particular dataset	Text	None	Mandatory	No	Depositor	Descriptive	Individual must have completed first questionnaire; must not have discontinued from the study; must not have lost pregnancy
00126	Collection mode	The method(s) used to collect the data in this dataset	Text	None	Mandatory	No	Depositor	Descriptive	Questionnaire
00127	Standardized Instruments Used	List previously validated instruments used in this dataset	Text	None	Recommended	No	Depositor	Descriptive	1. Perceived Stress Scale 2. Dyadic Adjustment Scale 3. Edinburgh Postnatal Depression Scale ...
00128	Pre-deposition cleaning	Description of any methods used to clean or preprocess the dataset prior to data deposition	Text	None	Recommended	No	Depositor	Descriptive	.
<i>SECTION C: ACCESS AND TERMS OF USE</i>									
00129	Availability	Whether or not the dataset is available for secondary re-use	Text	0= No; 1= Yes	Mandatory	No	Depositor	Descriptive	1
00130	Data security level	Minimum level of security necessary to access the dataset	Text	1 = Level 1 – access only through physical enclave; 2 = Level 2 – access using Virtual Research Environment; 3 = Level 3 – access via secure FTP 4 = Level 4 – Open Access	Mandatory	No	SAGE Staff	Descriptive	2

	Name of Field	Description	Type	Coding Scheme	Obligation	Repeatable	Created By	Intended Use	Example
00131	Vetting requirements	Any vetting requirements to export results from this dataset for reporting and publication that differs from the standard SAGE vetting requirements checklist. If none applicable, state “No special vetting requirements”	Text	None	Mandatory	No	SAGE Staff	Descriptive	No special vetting requirements

Element-Level Metadata

Element-level metadata are meant to capture information about an element (ie. variable) within a dataset. There are 14 elements in 2 sections.

Field ID	Name of Field	Description	Type	Coding Scheme	Obligation	Repeatable	Created By	Intended Use	Example
<i>SECTION A: ELEMENT DESCRIPTION</i>									
00132	Identifier	A unique identifier for the element	Text	None	Mandatory	No	Depositor	Descriptive	Q1MHW3
00133	Title	Descriptive title of the element. Can be a short form of the label if no formal title available	Text	None	Mandatory	No	Depositor	Descriptive	Mother's height
00134	Dataset association	Dataset to which this element belongs	Text	None	Mandatory	No	Depositor	Descriptive	Questionnaire 1 - Maternal
00135	Description	A description of the element. Can be the label, especially if no other description available	Text	None	Mandatory	No	Depositor	Descriptive	Self-reported height of mother, in inches
00136	Guide	Any additional information necessary to appropriately use the element, including important caveats.	Text	None	Mandatory if Applicable	No	Depositor	Descriptive	.
00137	Element Group	Group of elements to which this element belongs (eg. Item within a standardized scale, subject group)	Text	None	Recommended	No	Depositor	Descriptive	Item in Perceived Stress Scale
00138	Formatting	Standard (eg. ICD 10 diagnostic codes, CIHI depression rating scales) or custom labels that apply to element.	Text	None	Recommended	No	Depositor	Descriptive	.
00139	Formatting – link	If available, provide link to standard or custom formats	Linked File	None	Optional	No	Depositor	Descriptive	.
00140	Availability status	Does the element require review before release	Text	0=No; 1=Yes;	Mandatory	No	Depositor	Administrative	0
<i>SECTION B: ELEMENT CHARACTERISTICS</i>									
00141	Unit type	What is the unit of the data being measured?	Text	None	Recommended	No	Depositor	Descriptive	Inches
00142	Derived	Is the element computed from other elements?	Text	0=Not derived 1=Derived	Mandatory	No	Depositor	Descriptive	1
00143	Derivation procedure	If variable is derived, describe algorithm or methodology used to generate element	Text	None	Mandatory if Applicable	No	Depositor	Descriptive	(Q1MHW1*12) + Q1MHW2
00144	Imputation	If any missing values were imputed, describe procedure used to impute values to element	Text	None	Mandatory if Applicable	No	Depositor	Descriptive	.
00145	Unit of analysis	Whether element was collected about an individual, group, event or other IF DIFFERENT FROM DATASET LEVEL	Text	None	Mandatory if Applicable	No	Depositor	Descriptive	Neighbourhood level