ACT Data Harmonization Work Group

ACT Ontology and Data Dictionary
Version 1.4 *(updated March 1, 2017)*

Modification History

Version	Date	Modification / comment	Modified by
V1.0 (standard version naming convention)	3/2017	**All top level labels have been changed to move the ACT string. ACT is now a part of the version string.** **Demographics** Leading spaces were removed from age folders. **Diagnoses** Added integrated Diagnoses ICD10-ICD9 **Laboratory Tests** Fixed \ACT\Labs\LP14855-8\LP14400-3\LP43019-6\16128-1\ Hepatitis C virus Ab [Presence] in Serum **Medications** Created new hierarchy from RxNorm organized by PINs and MINs (multi-ingredient) where they exist. **Procedures** No change. **Visit Details** Leading spaces removed.	Shyam Visweswaran for DHWG
1.3	5/7/2015	• Added i2b2 basecodes	Shyam Visweswaran for DHWG
1.2	3/23/2015	• Changed title of document from "ACT Common Data Model" to "ACT Ontology and Data Dictionary" to reflect more accurately the nature of the document. • Demographic: SEX. Changed definition to "Sex" from "Administrative Sex" since this field reflects a blending of	Shyam Visweswaran for DHWG

		administrative gender and biological sex data. • Visit: VISIT_TYPE. Added "Emergency Department Admit to Inpatient Hospital Stay" and "Non-Acute Institutional Stay" to value set. • Medication: ORDER_TYPE AP=Ambulatory Prescribed, and ID=Inpatient Dispensed refer medication **orders** and not medication administration or medication billing, because data on orders will be available at all sites while administration and billing data may not be. Changed value set to A=Ambulatory Order and I=Inpatient Order. • In Medication and Laboratory Test tables, clarified that fields with "RAW" in the name are optional, and will not be used in queries.	
1.1	1/13/2015	First published version	Shyam Visweswaran for DHWG
1.0	11/25/2014	Initial version	Shyam Visweswaran for DHWG

Overview

The ACT Ontology and Data Dictionary document specifies the ontologies, the data domains, and the data elements that will be represented in Phase 1 of the ACT network.

Rows that have a grey background provide guidance for creating the SHRINE ontology file and are not to be implemented during the ETL process. Rows that have a clear background provide definitions and value sets for the common data elements and are to be implemented during the ETL process. Fields with "RAW" in the name may be optionally implemented by the ETL process, and will not be used in queries.

Missing or Unknown data values:

The ACT Ontology and Data Dictionary will use a single null value as a basis for representing missing or unknown values. Specifically, use NI=No Information which means:
1. A data field is not present in the source system.
2. A data field is present in the source system, but the source value is null or blank.
3. A data field is present in the source system, but the source value explicitly denotes an unknown value.
4. A data field is present in the source system, but the source value cannot be mapped to the common data model.

Demographic

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments
BIRTH_DATE	Date-time field	YYYY-MM-DD HH:MM:SS i2b2 basecode examples: 04 months old = DEM|AGE:0.4 55 year old = DEM|AGE:55	Date and time of birth. Current age (at time of query) in the SHRINE ontology is calculated from this. If times don't exist in the source data, set HH:MM:SS to 00:00:00.
SEX	TEXT(2)	A=Ambiguous F=Female M=Male O=Other NI=No information i2b2 basecodes: A = DEM|SEX:A F = DEM|SEX:F M = DEM|SEX:M	Sex. The "Ambiguous" category may be used for individuals who are physically undifferentiated from birth. The "Other" category may be used for individuals who are undergoing gender re-assignment.

		NI = DEM|SEX:NI O = DEM|SEX:O	
HISPANIC	TEXT(2)	Y=Yes N=No NI=No information i2b2 basecodes: Y = DEM|HISP:Y N = DEM|HISP:N NI = DEM|HISP:NI	A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. Uses "two question" approach.
RACE	TEXT(2)	01=American Indian or Alaska Native 02=Asian 03=Black or African American 04=Native Hawaiian or Other Pacific Islander 05=White 06=Multiple race NI=No information i2b2 basecodes: 01 = DEM|RACE:NA 02 = DEM|RACE:AS 03 = DEM|RACE:B 04 = DEM|RACE:H 05 = DEM|RACE:W 06 = DEM|RACE:M NI = DEM|RACE:NI	Use one or more race values per patient. American Indian or Alaska Native – A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment. Asian – A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. Black or African American – A person having origins in any of the black racial groups of Africa. Native Hawaiian or Other Pacific Islander – A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. White – A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

			Multiple - A person identifying themselves as more than one race. Uses "two question" approach. The value set is from U.S. Office of Management and Budget (OMB) standard, and is compatible with the 2010 U.S. Census.
VITAL_STATUS	TEXT(1)	D=Known Deceased i2b2 basecodes: D=DEM|VITAL STATUS:D	Note that NI is not allowed.
DEATH_DATE	Date-time field	YYYY-MM-DD HH:MM:SS	Date and time of death. Death date is not PHI. If times don't exist in the source data, set HH:MM:SS to 00:00:00.

Diagnosis

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments
DIAGNOSIS_CODING_SYSTEM	FIXED PERMISSABLE VALUE SET	ICD-9 ICD-10	Diagnosis coding system. In ACT Phase 1 we will use only ICD-9.
DIAGNOSIS_CODING_SYSTEM_VERSION	STRING	Example: Sixth Edition, 2008 (for ICD-9)	Diagnosis coding system version.
DIAGNOSIS_CODE	STRING		Diagnosis concept in coding system.
DIAGNOSIS_DATE	Date-time field	YYYY-MM-DD HH:MM:SS	Diagnosis date and time. If times don't exist in the source data, set HH:MM:SS to 00:00:00.
DIAGNOSIS_SOURCE (represented in the ontology as a modifier)	TEXT(2)	A=Admitting D=Discharge F=Final I=Interim NI=No information i2b2 basecodes: ACT|DIAG_SOURCE:A ACT|DIAG_SOURCE:D ACT|DIAG_SOURCE:F ACT|DIAG_SOURCE:I ACT|DIAG_SOURCE:NI	Classification of diagnosis source. We include these categories to allow some flexibility in implementation. The context is to capture available diagnoses recorded during a specific encounter. It is not necessary to populate interim diagnoses unless readily available. Ambulatory encounters would generally be expected to have a source of "Final."
DIAGNOSIS_PRIORITY (represented in the ontology as a modifier)	TEXT(2)	P=Principal S=Secondary NI=No information i2b2 basecodes: ACT|DIAG_PRIORITY:P ACT|DIAG_PRIORITY:S ACT|DIAG_PRIORITY:NI	Principal discharge diagnosis flag.

Procedure

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments
PROCEDURE_ CODING_SYSTEM	FIXED PERMISSABLE VALUE SET	ICD-9-CM ICD-10-PCS CPT-4 (i.e., HCPCS Level 1)	Procedure coding system. In ACT Phase 1 we will use only ICD-9-CM.
PROCEDURE_CODING_SYSTEM_VERSION	STRING	Example: Sixth Edition, 2008 (for ICD-9-CM)	Procedure coding system version.
PROCEDURE_CODE	STRING		Procedure concept in coding system.
PROCEDURE_DATE	Date-time field	YYYY-MM-DD HH:MM:SS	Procedure date and time. If times don't exist in the source data, set HH:MM:SS to 00:00:00.

Visit

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments
ADMIT_DATE	Date-time field	YYYY-MM-DD HH:MM:SS i2b2 basecode examples: Visit at age 04 months = DEM|AGE:0.4 Visit at age 55 years = DEM|AGE:55	Date and time of visit or admission. Age at visit field in SHRNE ontology is calculated from this. If times don't exist in the source data, set HH:MM:SS to 00:00:00.
DISCHARGE_DATE	Date-time field	YYYY-MM-DD HH:MM:SS	Date and time of discharge. Length of stay in SHRINE ontology is calculated from this. If times don't exist in the source data, set HH:MM:SS to 00:00:00.
VISIT_TYPE (represented in the ontology as a modifier)	TEXT(2)	AV=Ambulatory Visit ED=Emergency Department Visit EI=Emergency Department Visit To Inpatient IP=Inpatient Hospital Stay IS=Non-Acute Hospital Stay OA=Other Ambulatory Visit NI=No information i2b2 basecodes: ED = E EI = EI IP = I IS = NA OA = X NI = N	Visit type. Details of categorical definitions: Ambulatory Visit: Includes visits at outpatient clinics, physician offices, same day/ambulatory surgery centers, urgent care facilities, and other same-day ambulatory hospital encounters, but excludes emergency department encounters. Emergency Department (ED): Includes ED encounters that become inpatient stays (in which case inpatient stays would be a separate encounter). Excludes urgent care visits. ED claims should be pulled before hospitalization claims to ensure that ED with subsequent admission won't be rolled up in the hospital event.

			Emergency Department Admit to Inpatient Hospital Stay: Permissible substitution for preferred state of separate ED and IP records. Only for use with data sources where the individual records for ED and IP cannot be distinguished. Inpatient Hospital Stay: Includes all inpatient stays, including: same-day hospital discharges, hospital transfers, and acute hospital care where the discharge is after the admission date. Non-Acute Institutional Stay: Includes hospice, skilled nursing facility (SNF), rehab center, nursing home, residential, overnight non-hospital dialysis and other non-hospital stays. Other Ambulatory Visit: Includes other non-overnight AV encounters such as hospice visits, home health visits, skilled nursing facility visits, other non-hospital visits, as well as telemedicine, telephone and email consultations. May also include "lab only" visits (when a lab is ordered outside of a patient visit), "pharmacy only" (e.g., when a patient has a refill ordered without a face-to-face visit), "imaging only", etc.

Medication

Notes:
- Each entry in this table represents a medication order, not medication administration or medication billing.
- Fields with "RAW" in the name are optional, and will not be used in queries.

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments
MEDICATION_CODING_SYSTEM	FIXED PERMISSABLE VALUE SET	RxNorm	Medication coding system.
MEDICATION_CODING_SYSTEM_VERSION	STRING	Example: RxNorm 01-Dec-2014;17-Dec-2014	Medication coding system version.
MEDICATION_CLASSIFICATION_SYSTEM	FIXED PERMISSABLE VALUE SET	"Products by VA Class" classification from the National Drug File - Reference Terminology (NDF-RT).	Medication classification system.
MEDICATION_CLASSIFICATION_SYSTEM_VERSION	STRING	Example: NDF-RT December 2014 Version	Medication classification system version.
MEDICATION_CODE	STRING	RxNorm RxCUI	Medication concept in coding system. For ACT, map drugs to RxNorm's concepts of Semantic Clinical Drug (SCD), Semantic Branded Drug (SBD), Generic Pack (GPCK), or Branded Pack (BPCK). These concepts contain drug name, strength, form, and route of administration. The ACT ontology currently does not have complete coverage for all medications; however, map all of the available medications to SCD or SBD. If unable to map to SCD/SBD map to IN.

ORDER_DATE	Date-time field	YYYY-MM-DD HH:MM:SS	Order date and time. If times don't exist in the source data, set HH:MM:SS to 00:00:00.
ORDER_TYPE (represented in the ontology as a modifier)	TEXT(2)	I=Inpatient Order A=Ambulatory Order NI=No information i2b2 basecodes: I = ACT|MED_ORDER_TYPE:I A = ACT|MED_ORDER_TYPE:A NI = ACT|MED_ORDER_TYPE:NI	Location where medication was ordered.
RAW_ MEDICATION_CODE	TEXT(x)		Optional field for originating local code related to a medication order.
RAW_ MEDICATION_NAME	TEXT(x)		Optional field for originating local medication name.
RAW_ MEDICATION_STRENGTH	TEXT(x)		Optional field for originating local medication strength.
RAW_ MEDICATION_FORM	TEXT(x)		Optional field for originating local medication form.
RAW_ MEDICATION_ROUTE	TEXT(x)		Optional field for originating local medication route.

Laboratory Test

Note:
- Fields with "RAW" in the name are optional, and will not be used in queries.

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments
LAB_ CODING_SYSTEM	FIXED PERMISSABLE VALUE SET	Example: LOINC (for now)	Laboratory test coding system.
LAB_CODING_SYSTEM_VERSION	STRING	Example: LOINC 2.48;2014-06-27	Laboratory test coding system version.
LAB_ CLASSIFICATION_SYSTEM	FIXED PERMISSABLE VALUE SET	Example: LOINC multi-axial codes (codes that start with LP) (for now)	Laboratory test classification system.
LAB_ CLASSIFICATION _SYSTEM_VERSION	STRING	Example: LOINC 2.48;2014-06-27	Laboratory test classification system version.
LAB_CODE	FIXED PERMISSABLE VALUE SET	LOINC laboratory test code.	Laboratory test concept in coding system. Only a subset of labs are to be mapped for ACT Phase 1. See ACT SHRINE Query Ontology document for details.
RESULT_LOCATION (represented in the ontology as a modifier)	TEXT(2)	L=Laboratory P=Point of Care NI=No information i2b2 basecodes: L = ACT|LAB_LOC:L P = ACT|LAB_LOC:P NI = ACT|LAB_LOC:NI	Location of the test result. Point of Care locations may include anticoagulation clinic, newborn nursery, finger stick in provider office, or home. The default value is 'L' unless the result is Point of Care. There should not be any missing values.
SPECIMEN_DATE	Date-time field	YYYY-MM-DD HH:MM:SS	Date and time specimen was collected. If times don't exist in the source data, set HH:MM:SS to 00:00:00.
RESULT_QUALITATIVE	TEXT(10)	BORDERLINE POSITIVE NEGATIVE UNDETERMINED NI=No information	Standardized result for qualitative results. This variable should be NI for quantitative results. This filed will be sued for other

			permissible value sets e.g., color of urine.
RESULT_NUMERICAL	INTEGER		Standardized/converted result for quantitative results. This variable should be left blank for qualitative results.
RESULT_MODIFIER	TEXT	EQ=Equal GE=Greater than or equal to GT=Greater than LE=Less than or equal to LT=Less than TX=Text NI=No information	Modifier for result values. Any symbols in the RAW_RESULT value should be reflected in the RESULT_MODIFIER variable. For example, if the original source data value is "<=200" then RAW_RESULT=200 and RESULT_MODIFIER=LE. If the original source data value is text then RESULT_MODIFIER=TX. If the original source data value is a numeric value then RESULT_MODIFIER=EQ.
RESULT_UNIT	FIXED PERMISSABLE VALUE SET		Converted/standardized units for the result. Use Unified Code for Units of Measure (UCUM). The ACT SHRINE Query Ontology specifies UCUM units for each lab.
ABNORMAL_RESULT_INDICATOR	TEXT(2)	AB=Abnormal AH=Abnormally high AL=Abnormally low CH=Critically high CL=Critically low CR=Critical IN=Inconclusive NL=Normal NI=No information	Abnormal result indicator. This value comes from the source data; do not apply logic to create it.
RAW_LAB_CODE	TEXT(x)		Local code related to an individual lab test.
RAW_PANEL	TEXT(x)		Local code related to a battery or panel of lab tests.
RAW_RESULT	TEXT(x)		The original test result value as seen in your source data. Values may include a decimal

			point, a sign or text (e.g., POSITIVE, NEGATIVE, DETECTED). The symbols >, <, >=, <= should be removed from the value and stored in the Modifier variable instead.
RAW_UNIT	TEXT(x)		Original units for the result in your source data.
RAW_ORDER_DEPT	TEXT(x)		Local code for ordering provider department.
RAW_FACILITY_CODE	TEXT(x)		Local facility code that identifies the hospital or clinic. Taken from facility claims.
RAW_NORMAL_RANGE_LOW	TEXT(10)		Lower bound of the normal range assigned by the laboratory. Value should only contain the value of the lower bound. The symbols >, <, >=, <= should be removed. For example, if the normal range for a test is >100 and <300, then "100" should be entered.
RAW_MODIFIER_LOW	TEXT(2)	EQ=Equal GE=Greater than or equal to GT=Greater than NO=No lower limit NI=No information	Modifier for RAW_NORMAL_RANGE_LOW values. For numeric results one of the following needs to be true: 1) Both RAW_MODIFIER_LOW and RAW_MODIFIER_HIGH contain EQ (e.g., normal values fall in the range 3-10) 2) RAW_MODIFIER_LOW contains GT or GE and RAW_MODIFIER_HIGH contains NO (e.g., normal values are >3 with no upper boundary) 3) RAW_MODIFIER_HIGH contains LT or LE and RAW_MODIFIER_LOW contains NO (e.g., normal

			values are <=10 with no lower boundary)
RAW_NORMAL_RANGE_ HIGH	TEXT(11)		Upper bound of the normal range assigned by the laboratory. Value should only contain the value of the upper bound. The symbols >, <, >=, <= should be removed. For example, if the normal range for a test is >100 and <300, then "300" should be entered.
RAW_MODIFIER_HIGH	TEXT(2)	EQ=Equal LE=Less than or equal to LT=Less than NO=No higher limit NI=No information	Modifier for RAW_NORMAL_RANGE_HIG H values. For numeric results one of the following needs to be true: 1) Both RAW_MODIFIER_LOW and RAW_MODIFIER_HIGH contain EQ (e.g., normal values fall in the range 3-10) 2) RAW_MODIFIER_LOW contains GT or GE and MODIFIER_HIGH contains NO (e.g., normal values are >3 with no upper boundary) 3) RAW_MODIFIER_HIGH contains LT or LE and MODIFIER_LOW contains NO (e.g., normal values are <=10 with no lower boundary)