

HITSP Anonymize Long Term and Post Acute Care Assessment Data Component

HITSP/C165



Healthcare Information Technology Standards Panel

Submitted to:

Healthcare Information Technology Standards Panel

Submitted by:

Security, Privacy and Infrastructure Domain Technical Committee



DOCUMENT CHANGE HISTORY

Version Number	Description of Change	Name of Author	Date Published
0.0.1	Review Copy	Security, Privacy and Infrastructure Domain Technical Committee	January 31, 2010



TABLE OF CONTENTS

1.0	INTRODUCTION.....	5
1.1	Overview.....	5
1.2	Copyright Permissions.....	5
1.3	Reference Documents.....	5
1.4	Conformance	5
1.4.1	Conformance Criteria	5
1.4.2	Conformance Scoping, Subsetting and Options	6
2.0	COMPONENT DEFINITION.....	7
2.1	Context Overview	7
2.1.1	Component Dependencies	8
2.2	Rules for Implementing.....	8
2.2.1	Anonymity Levels	8
2.2.1.1	Level 1 Anonymity: Removal of Clearly Identifying Data	8
2.2.1.2	Level 2 Anonymity: Static Model Based Re-identification Risk Analysis	9
2.2.1.3	Level 3 Anonymity: Routine Resource Risk Analysis	10
2.2.2	Data Mapping	10
2.2.2.1	Level 1 Anonymity Considerations.....	10
2.2.2.2	Level 2 Anonymity Considerations.....	12
2.3	Standards	13
2.3.1	Regulatory Guidance.....	13
2.3.2	Selected Standards	13
2.3.3	Informative Reference Standards.....	13
3.0	APPENDIX	14
4.0	DOCUMENT UPDATES	15
4.1	January 31, 2010.....	15



FIGURES AND TABLES

Table 1-1 Reference Documents	5
Table 2-1 Component Dependencies	8
Table 2-2 Data Mapping – MDS 3.0	10
Table 2-3 Data Mapping – OASIS-C.....	11
Table 2-4 Regulatory Guidance	13
Table 2-5 Selected Standards	13
Table 2-6 Informative Reference Standards	13



1.0 INTRODUCTION

1.1 OVERVIEW

Anonymization, according to the International Organization for Standardization (ISO), is the process that removes the association between the identifying data set and the data subject. The Healthcare Information Technology Standards Panel (HITSP) Anonymize Long Term and Post Acute Care Data Component provides specific instruction for anonymizing data that are prepared for repurposing data created as part of routine clinical care delivery. This construct defines the Component specification that provides the ability to anonymize patient identifiable information for Long Term Care reporting data.

Anonymization cannot be guaranteed by the use of this construct, and therefore a comprehensive risk assessment should be conducted in the implementation environment.

1.2 COPYRIGHT PERMISSIONS

COPYRIGHT NOTICE

© 2010 ANSI. This material may be copied without permission from ANSI only if and to the extent that the text is not altered in any fashion and ANSI's copyright is clearly noted.

1.3 REFERENCE DOCUMENTS

A list of key reference documents and background material is provided in the table below. HITSP-maintained reference documents can be retrieved from the [HITSP Web Site](#).

Table 1-1 Reference Documents

Reference Document	Document Description
HITSP Acronyms List	Lists and defines the acronyms used in this document
HITSP Glossary	Provides definitions for relevant terms used by HITSP documents
TN900 – Security and Privacy	TN900 is a reference document that provides the overall context for use of the HITSP Security and Privacy constructs
TN901 – Clinical Documents	TN901 is a reference document that provides the overall context for use of the HITSP Care Management and Health Records constructs
TN903 – Data Architecture	TN903 is a reference document that provides the overall context for use of the HITSP Data Architecture constructs
TN904 – Harmonization Framework and Exchange Architecture	TN904 is a reference document that provides the overall context for use of the HITSP Harmonization Framework and Exchange Architecture

1.4 CONFORMANCE

This section describes the conformance criteria, which are objective statements of requirements that can be used to determine if a specific behavior, function, interface, or code set has been implemented correctly.

1.4.1 CONFORMANCE CRITERIA

In order to claim conformance to this construct specification, an implementation must satisfy all the requirements and mandatory statements listed in this specification, the associated HITSP Interoperability Specification or Capability, its associated construct specifications, as well as conformance criteria from the selected base and composite standards. A conformant system must also implement all of the required



interfaces within the scope, subset or implementation option that is selected from the associated Interoperability Specification.

Claims of conformance may only be made for the overall HITSP Interoperability Specification or Capability with which this construct is associated.

1.4.2 CONFORMANCE SCOPING, SUBSETTING AND OPTIONS

A HITSP Interoperability Specification or Capability must be implemented in its entirety for an implementation to claim conformance to the specification. HITSP may define the permissibility for interface scoping, subsetting or implementation options by which the specification may be implemented in a limited manner. Such scoping, subsetting and options may extend to associated constructs, such as this construct. This construct must implement all requirements within the selected scope, subset or options as defined in the associated Interoperability Specification or Capability to claim conformance.



2.0 COMPONENT DEFINITION

2.1 CONTEXT OVERVIEW

The Long Term Care and Post Acute Care Assessment (LTPAC) Use Case is focused on the electronic exchange of Long Term Care information among clinicians and others for purposes of care for Long Term Care Patients.

This Component supports communication of de-identified LTPAC data to Population Health and others in support of activities and programs related to LTPAC populations. Population Health may benefit from the ability to receive de-identified LTPAC Assessment information to support population health or research activities for medical conditions and additional assessment information available from LTPAC Assessments.

Note: Assessments contain information such as immunizations, broad demographics, etc., that can be useful to Public Health.

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) regulation in 45 CFR 164.512(a) states:

“A covered entity may use or disclose protected health information to the extent that such use or disclosure is required by law and the use or disclosure complies with and is limited to the relevant requirements of such law.”

45 CFR 164.512(b) states:

“A covered entity may disclose protected health information for the public health activities and purposes described in this paragraph to: (i) A public health authority that is authorized by law to collect or receive such information for the purpose of preventing or controlling disease, injury, or disability, including, but not limited to, the reporting of disease, injury, vital events such as birth or death, and the conduct of public health surveillance, public health investigations, and public health interventions; or, at the direction of a public health authority, to an official of a foreign government agency that is acting in collaboration with a public health authority.”

HITSP interprets the rule to permit covered entities to disclose protected health information without authorization for specified public health purposes. However, this permissiveness is not a request to any and all data. In practice, LTPAC supports such data requests with rational supporting justification. This commonly takes the form of state or local legislation declaring the types of data, or specific data elements required for public health. HITSP recommends that all implementations undergo legal review to ensure compliance with local, state, and federal regulations. HITSP supports further harmonization of policy and practices for more uniform public health data exchange.

HITSP supports the use of anonymization and pseudonymization approaches to protect individual privacy and confidentiality. This Component specifies anonymization protections for such data.

Under 45 CFR 164.502(d), HIPAA defines 18 data elements that under a Safe Harbor approach must be removed from personal health records in order for those records to be considered anonymized. There are some demographic data elements of interest that need to be retained in order to accurately evaluate the data. This Component specifies removal and aggregation requirements for data variables used and disclosed by LTPAC entities.

The selected standard is the ISO Health informatics – Pseudonymisation, Technical Specification #25237 (ISO TS25237). This standard defines 3 levels of anonymization, with specific requirements for anonymization at each one of those anonymization levels. These requirements are described in Section 2.2.



2.1.1 COMPONENT DEPENDENCIES

Table 2-1 Component Dependencies

Standard/HITSP Component	Depends On (Name of standard/HITSP Component that it depends on)	Dependency Type (Pre-condition, Post-condition, General)	Purpose (Reason for this dependency)
HITSP/C165 – Anonymize Long Term and Post Acute Care Assessment Data	HITSP/C154 – Data Dictionary	General	Identification of Data Elements constrained within this Component

2.2 RULES FOR IMPLEMENTING

2.2.1 ANONYMITY LEVELS

The ISO Pseudonymisation (ISO TS25237) specification defines the following level concepts with respect to anonymity.

- Level 1 Anonymity: Removal of Clearly Identifying Data
- Level 2 Anonymity: Static Model Based Re-identification Risk Analysis
- Level 3 Anonymity: Routine Resource Risk Analysis

2.2.1.1 LEVEL 1 ANONYMITY: REMOVAL OF CLEARLY IDENTIFYING DATA

A first, intuitive level of anonymity can be achieved by applying rules of thumb. This method is usually implicitly understood when de-identifying data are discussed. In many contexts, this first level of anonymity may provide a sufficient guarantee.

As an example of Level 1 Anonymity, the HIPAA rule is given. The HIPAA rule requires that for data to be considered de-identified, the following elements should be removed:

- Names (individual, employer, relatives, etc.)
- Address (street, city, county, precinct, zip code – initial 3 digits if geographic unit contains less than 20,000 people, or any other geographical codes)
- Telephone and Fax numbers
- Social Security numbers
- Dates (except for years)
 - Birth date
 - Admission date
 - Discharge date
 - Date of death
 - Ages >89 and all elements of dates indicative of such age (except that such age and elements may be aggregated into a category “Age >90”)
 - E-mail addresses
- Health Plan Beneficiary numbers
- Account numbers
- Certificate/license numbers
- Vehicle Identifiers and Serial Numbers (e.g., VINs, license plate numbers)
- Device Identifiers and Serial Numbers
- Web Universal Resource Locators (URLs)
- Internet Protocol (IP) address numbers
- Biometric Identifiers (e.g., finger or voice prints)
- Full face photographic images and any comparable images
- Any other unique identifying number, characteristic, or code



2.2.1.2 LEVEL 2 ANONYMITY: STATIC MODEL BASED RE-IDENTIFICATION RISK ANALYSIS

The second level of anonymity takes into account the global data model and the data flows inside the model. This level includes a static risk analysis that checks for re-identification vulnerabilities by different interfaces. This level may for example include the removal of absolute time references. A reference time marker “T” is defined as the admission of a patient for an episode of care and other events; discharge is expressed with reference to this time marker.

Level 2 Anonymity Issues with Free-Form Text

Free text and privacy are not very compatible since the notion of “free” denotes the opposite of structured. In Information Technology (IT) terminology, the notions of “data” and “information” are treated separately. Structured data give some indication of what information can be expected where. It is then up to re-identification risk analysis to make assumptions about what can lead to (unacceptable) identification risks, ranging from simple rules of thumb as specified in HIPAA, to analysis of populated databases and inference deductions.

In “free text”, as opposed to “structured,” there is no way to begin automated analysis for privacy purposes with a guaranteed outcome (and the derived liabilities). For example, the presence and position of an information item in a free text document may not be predictable but when it is present, it can be deduced from a pattern (e.g., a sentence like ‘the patient had complaints about’ or ‘patient <name> was discharged at ...’). Simple pattern parsing or enhanced Natural Language Processing (NLP) can deduce structure in those cases, but perhaps not for the whole text. The notion “free” is more connected to unpredictability of presence or position of information elements. Structure is obtained by the ability to extract data elements either through fixed position, delimiters or tags. Even then, a user may input data elements (e.g., put a patient number where a diagnosis should be put), but the certainty about the content is higher in structured documents.

There can be a discussion on how unstructured “free text” is. Policies could define some rules (e.g., define that the free text part shall not contain directly identifiable information such as patient numbers, names, or CFR rule of thumb items such as defined in HIPAA). Parsing and NLP could be applied to separate directly identifying items (e.g., numbers with a certain length, structure or preamble). In some cases, the free text originates from structured text (e.g., an automated letter of discharge from a hospital generated from the hospital’s Health Information System). This makes it easier for the parsing or NLP.

Ultimately, the primary de-identification decision is to:

- Single out what, according to your policy and desired anonymity level, is identifiable information
- Delete what you don’t need
- Keep together (in the payload) what is considered according to the policy as non-identifiable

A hospital policy could specify that investigators cannot put identifiable information into the free text component and define what is meant by identifiable. From a privacy point of view this turns it into structured data with the payload containing free text. The liability for privacy violations is shifted towards the editor of the free text to stick to the agreed policies. From a privacy point of view the baseline on deciding if text is free is the following:

- Parts (possibly) containing identification are known
- Parts denoted as non-identifying should at least not contain nominative information
- Hybrid situations are possible (e.g., the part with identification is structured but the rest unstructured)



2.2.1.3 LEVEL 3 ANONYMITY: ROUTINE RESOURCE RISK ANALYSIS

An anonymized resource used for data mining must undergo a routine statistical evaluation for re-identification risks associated with the populated resource. Such risk analysis entails assessments of outliers and analytical linking with external information resources.

2.2.2 DATA MAPPING

Different jurisdictions and stakeholders will have different requirements and agreements that may not require full anonymization of these data elements (see Context Overview Section 2.1). In consideration of the HIPAA Rules and ISO Pseudonymisation (ISO TS25237), the following sections describe anonymization requirements associated with collecting and retaining an information repository for public health case reporting.

2.2.2.1 LEVEL 1 ANONYMITY CONSIDERATIONS

To be compliant with full de-identification, all patient identifying information specified by HIPAA must be removed from the message or document to be submitted for repurposing. The following anonymity rules apply to the data variables specified below, as described in the Interoperability Specification or Capability calling this construct.

Note that it is anticipated that facilities will act to shield identities by using contact details (phone number, address, contact person, etc.) that do not identify the facility.

Table 2-2 and Table 2-3 describe the data mapping to be anonymized, as specified in MDS 3.0 and OASIS-C.

Note: In a future maintenance release of this construct and all HITSP Anonymization constructs, data mapping will reference HITSP/C154 Data Dictionary.

Table 2-2 Data Mapping – MDS 3.0

Item ID	Identification Information	"Common" Item Name	Anonymity Action
A0100A		National Provider Identifier	PASS
A0100B		CMS Certification Number (CCN)	PASS
A0100C		State Provider Number	PASS
A0200	Type of Provider	Nursing home (SNF/NF)	PASS
A0200	Type of Provider	Swing bed	PASS
A0310B	PPS Assessment--to establish Medicare A payment rate		PASS [C001]
A0310C (1,2 or 3)	PPS Other Medicare Required		PASS [C001]
A0300D 1	Is this a swing bed clinical change Assessment?		PASS [C001]
A0410 (1,2 or 3)	Submission Requirement		PASS [C001]
A0500A		First Name	BLIND
A0500B		Middle Initial	BLIND
A0500C		Last Name	BLIND
A0500D		Suffix	BLIND
A0600A	Social Security Number	Resident Social Security Number	BLIND
A0600B	Medicare Number (Or Comp.Rr Ins #)	Resident Medicare Number (Or Comparable Railroad Insurance Number)	BLIND
A0700	Medicaid Number	Any Response	BLIND
A0800	Gender--Administrative		BLIND
A0900	Resident birthdate		PASS [C002]



Item ID	Identification Information	"Common" Item Name	Anonymity Action
A1000	Race/Ethnicity	CDC Race and Ethnicity Code Set, value set OID 2.16CodeSystem OID 2.16.840.1.113883.6.238	PASS [C001]
A1100B		Specify preferred language	BLIND
A1200	Marital Status	Never married	BLIND
A1300A		Facility medical record number	BLIND
A1300B		Resident room number	BLIND
A1300C		Name by which resident prefers to be addressed	BLIND
A1300D		Lifetime occupation(s) - put "/" between two occupations--TEXT FIELD	BLIND
A1500		Not a Medicaid certified unit	PASS [C001]
A1600		Entry Date (date of this admission/reentry into the facility)	BLIND
A2400B		Start date of most recent Medicare stay	BLIND
A2400C		Medicare stay is ongoing	BLIND

[C001] – Pass through only if 2 or more meet criteria

[C002] – Pass through MM/YYYY if <89 years old, otherwise blind

Table 2-3 Data Mapping – OASIS-C

Item Number	Identification Information	"Common" Item Name	Anonymity Rule
M0040	M0040_PAT_FNAME	PATIENT FIRST NAME	BLIND
M0040	M0040_PAT_MI	PATIENT MIDDLE INITIAL	BLIND
M0040	M0040_PAT_LNAME	PATIENT LAST NAME	BLIND
M0040	M0040_PAT_SUFFIX	PATIENT NAME SUFFIX	BLIND
M0050	M0050_PAT_ST	PATIENT STATE	BLIND
M0060	M0060_PAT_ZIP	PATIENT ZIP	pass--1st 3 digits of zip only
M0063	M0063_MEDICARE_NUM	MEDICARE #	BLIND
M0064	M0064_SSN	SSN	BLIND
M0065	M0065_MEDICAID_NUM	MEDICAID #	BLIND
M0066	M0066_PAT_BIRTH_DT	BIRTHDATE--MMDDYYYY	PASS [C002]
M0069	M0069_PAT_GENDER	GENDER-1-M/2-F--administrative	PASS
M0016	M0016_Branch_ID	Home Health Branch NPI	PASS
M0018	M0018_PHYSICIAN_ID	PHYSICIAN NPI	PASS
M0140	Race/Ethnicity		PASS
M0150	PAYMENT SOURCE	PAYMENT SOURCE	PASS
	HX AND DX		
M1000	D/C FROM IN LAST 14 DAYS	From which of the following Inpatient Facilities was the patient discharged during the past 14 days? (Mark all that apply.)	PASS
M1000	D/C FROM IN LAST 14 DAYS	1 - Long-term nursing facility (NF)	PASS



Item Number	Identification Information	"Common" Item Name	Anonymity Rule
M1000	D/C FROM IN LAST 14 DAYS	2 - Skilled nursing facility (SNF/TCU)	PASS
M1000	D/C FROM IN LAST 14 DAYS	3 - Short-stay acute hospital (IPPS)	PASS
M1000	D/C FROM IN LAST 14 DAYS	4 - Long-term care hospital (LTPACH)	PASS
M1000	D/C FROM IN LAST 14 DAYS	5 - Inpatient rehabilitation hospital or unit (IRF)	PASS
M1000	D/C FROM IN LAST 14 DAYS	6 - Psychiatric hospital or unit	PASS
M1000	D/C FROM IN LAST 14 DAYS	7 - Other (specify) (TEXT FIELD)	BLIND
M1005	M1005_INP_DISCHARGE_DT	INPATIENT D/C DATE--MM/DD/YYYY	PASS [C003]
M0102	M0102_PHYSN_ORDRD_SOCROC_DT	Date of Physician-ordered Start of Care (Resumption of Care): If the physician indicated a specific start of HOME HEALTH care mm/dd/yyyy	PASS
M0104	M0104_PHYSN_RFRL_DT	Date of Referral: Indicate the date that the written or verbal referral for initiation or resumption of care was received by the HHA. Mm/dd/yyyy	PASS
M0090	M0090_INFO_COMPLETED_DT	INFORMATION COMPLETION DATE--ASSESSMENT COMPLETION DATE mm/dd/yyyy Although USUALLY completed on the first date of services, it can be the second.	PASS

[C001] – Pass through only if 2 or more meet criteria

[C002] – Pass through MM/YYYY if <89 years old, otherwise blind

[C003] –Pass through DD/MM/YYYY unless D/C = 5 (M1000)

2.2.2.2 LEVEL 2 ANONYMITY CONSIDERATIONS

This section describes the Level 2 Anonymity considerations that pertain to the data elements.

Inference Risk Mitigations:

Freeform data poses a privacy and confidentiality risk because freeform text can contain identifiable information. If one would like to get value out of that data, then they need to develop methods to extract codified information. To be compliant with full de-identification, this approach should be applied to freeform text within the message or document to be submitted for repurposing. No stipulation is made in this specification as to the algorithms or process by which the codification is accomplished. Future specifications may address this area further, but until then, the approach is left to the implementer.

Because of the re-identification risks within the Data Set identified in Section 2.2.2 of this document, the information resource must have access restricted to authorized persons contractually bound or otherwise bound (and subject to sanction) to use the resource for specified purposes. This Component specification recommends that access be restricted to authorized public health authorities and infection control clinical staff associated with healthcare provider organizations.



2.3 STANDARDS

2.3.1 REGULATORY GUIDANCE

Table 2-4 Regulatory Guidance

Regulation	Description
Health Insurance Portability and Accountability Act (HIPAA) Code of Federal Regulations (CFR) Title 45, Part 164, Section 502(d) (CFR§164.502(d)) Uses and disclosures of protected health information: general rules	This is a specific reference to 45 CFR 164.502(d) which specifies the general rules for uses and disclosures of de-identified protected health information

2.3.2 SELECTED STANDARDS

Table 2-5 Selected Standards

Standard	Description
International Organization for Standardization (ISO) Health Informatics -- Pseudonymisation, Technical Specification # 25237 (ISO TS25237)	Health Informatics – Pseudonymisation. Approved as a Technical Specification March, 2007. For more information visit www.iso.org

2.3.3 INFORMATIVE REFERENCE STANDARDS

Table 2-6 Informative Reference Standards

Standard	Reason for Use
Digital Imaging and Communications in Medicine (DICOM) Attribute Level Confidentiality Supplement: # 55	Adds a mechanism for selective protection of individual attributes within arbitrary DICOM service-object pair (SOP) instances. It may be used to achieve protection of identifying information, e.g., a reversible anonymization or pseudonymization of DICOM SOP instances while continuing to use unmodified lower level message and protocol services for network transfer, storage, and media exchange of composite image information objects. For more information visit medical.nema.org



3.0 APPENDIX

The following sections include relevant materials referenced throughout this document.

No additional information at this time.



4.0 DOCUMENT UPDATES

The following sections provide the details of updates made to this document.

4.1 JANUARY 31, 2010

No changes. This is the first published version of the document.

