

HITSP Anonymize Newborn Screening Component

HITSP/C164



Healthcare Information Technology Standards Panel

Submitted to:

Healthcare Information Technology Standards Panel

Submitted by:

**Security, Privacy and Infrastructure Domain Technical Committee
(Formerly Security and Privacy 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	November 9, 2009
0.0.2	Review Copy	Security, Privacy and Infrastructure Domain Technical Committee	January 18, 2010
1.0	Released for Implementation	Security, Privacy and Infrastructure Domain Technical Committee	January 25, 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 Constraints.....	8
2.1.2	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.....	11
2.3	Standards	12
2.3.1	Regulatory Guidance.....	12
2.3.2	Selected Standards	12
2.3.3	Informative Reference Standards.....	12
3.0	APPENDIX	13
4.0	DOCUMENT UPDATES	14
4.1	November 9, 2009	14
4.2	January 18, 2010.....	14
4.3	January 25, 2010.....	14



FIGURES AND TABLES

Table 1-1 Reference Documents	5
Table 2-1 Component Constraints	8
Table 2-2 Component Dependencies	8
Table 2-3 Data Mapping Level 1 Patient Data Elements – Pediatric Demographic Data	10
Table 2-4 Data Mapping Level 1 Patient Data Elements – Results Data	11
Table 2-5 Regulatory Guidance	12
Table 2-6 Selected Standards	12
Table 2-7 Informative Reference Standards	12



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 Newborn Screening 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 Newborn Screening 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-managed reference documents may be retrieved from the www.hitsp.org 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

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

This construct provides guidance for anonymization and should be implemented with consideration of risk assessment results in the intended operating environment. This construct is intended specifically for the use of anonymizing Newborn Screening Data, and should not be reused for any other purpose.

The Newborn Screening (NBS) Use Case is focused on the electronic exchange of information related to newborn screening among ordering clinicians, pediatric clinicians, consumers, Public Health, testing laboratories, and audiology service providers. Newborn screening reporting and information exchanges may also include individual case reporting to Public Health, appropriate registries, and local health service providers.

The scope of this construct is focused on the ability to report newborn screening information between:

- Providers and Public Health agencies
- Public Health agencies

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, public health 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.

Disclosure of patient identifiable data to public health authorities in the context of reportable conditions monitoring is routine; this disclosure is based upon the need to monitor and manage known public health threats. Public health systems collect a broad variety of healthcare data that may go beyond capturing data to support assessment of known threats. As such, HITSP supports the use of anonymization and pseudonymization approaches to protect individual privacy and confidentiality. This Component specifies anonymization protections for such data collection.

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. The Use Case has defined some demographic data elements of interest that need to be retained in order to accurately evaluate the data for public health case reporting. This Component specifies removal and aggregation requirements for data variables submitted to a Public Health Information System.



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 CONSTRAINTS

The use of this construct assumes that all policy agreements and regulatory requirements applicable to the purpose for which the construct is being used are adhered to by the parties exchanging the information. In the absence of regulatory requirements, the use of this construct will be possible because of an agreement between the exchange parties.

Table 2-1 Component Constraints

Constraint	Constraint Section
With the exception of the data variables described in Table 2-3 below, all identifiers from the list of identifiable data variables defined by HIPAA that must be removed to accomplish de-identification are expected to be removed from the documents and messages that are communicated to the Public Health System	N/A
This construct does not address anonymization and/or pseudonymization issues associated with adoption of children	N/A
This construct does not address additional clinical elements (e.g. HITSP/C32, Medications, Allergies, Problems, Family History) which may be communicated along with the results	N/A

2.1.2 COMPONENT DEPENDENCIES

Table 2-2 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/C164 - Anonymize Newborn Screening	HITSP/C154 –Data Dictionary	General	Data Dictionary Definitions

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 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 2-3 and Table 2-4 describe the data mapping to be anonymized. The term "Public Health to Public Health" includes State to Federal reporting,

Table 2-3 Data Mapping Level 1 Patient Data Elements – Pediatric Demographic Data

Data Element	Anonymization Requirements Provider to Public Health	Anonymization Requirements Public Health to Public Health
Patient Name: First, Middle, Last	Blind	Blind
Patient Alias Name: First, Middle, Last	Blind	Blind
Patient Address	Allow first 3 digits of zip code	Allow first 3 digits of zip code
Patient Phone Number	Blind	Blind
Patient Identifier	Pseudonym	Blind



Data Element	Anonymization Requirements Provider to Public Health	Anonymization Requirements Public Health to Public Health
Patient Birth Date	Pass Through	Month/Year
Patient Sex	Pass Through	Pass Through
Patient Race	Pass Through	Pass Through
Patient Ethnicity	Pass Through	Pass Through
Patient Primary Language	Blind	Blind
Patient Multiple Birth Indicator	Blind	Blind
Patient Multiple Birth Order	Blind	Blind
Patient Birth Registration Number	Blind	Blind
Patient Birth State/Country	Pass through	Pass through
Patient Birthing Facility	Pass Through	Pass Through
Mother's Name: First, Middle, Last	Blind	Blind
Mother's Maiden Name	Blind	Blind
Mother's SSN	Blind	Blind
Father's Name: First, Middle, Last	Blind	Blind
Father's SSN	Blind	Blind
Insurance Plan	Public/Private (need constrained value set)	Blind
Insurance Company	Blind	Blind
Immunization Services Funding Eligibility	Blind	Blind
Next of Kin Relationship	Blind	Blind
Next of Kin Address	Blind	Blind
Next of Kin Telephone	Blind	Blind
Next of Kin DOB	Blind	Blind
Last Update Time/Date	Pass Through	Pass Through
Last Update Facility	Pass Through	Pass Through

Table 2-4 Data Mapping Level 1 Patient Data Elements – Results Data

Data Element	Anonymization Requirements Provider to Public Health	Anonymization Requirements Public Health to Public Health
15.01 Resulted test	Pass Through	Pass Through
15.05 Result value	Pass Through	Pass Through
15.05 Result unit	Pass Through	Pass Through
15.02 Report date/time	Pass Through	Pass Through
15.04 Result status	Pass Through	Pass Through
15.06 Test interpretation	Blind if free text, Pass through if coded field	Blind if free text, Pass through if coded field

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 2-5 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 2-6 Selected Standards

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

2.3.3 INFORMATIVE REFERENCE STANDARDS

Table 2-7 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.

RELEASED FOR IMPLEMENTATION



4.0 DOCUMENT UPDATES

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

4.1 NOVEMBER 9, 2009

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

4.2 JANUARY 18, 2010

This document has been updated to the revised Component template Version 2.6.

4.3 JANUARY 25, 2010

Upon approval by the HITSP Panel on January 25, 2010, this document is now Released for Implementation.

