# Digital Imaging and Communications in Medicine (DICOM)

*Supplement 55: Attribute Level Confidentiality (including De-identification)*

**DICOM Standards Committee, Working Group 14 Security**

1300 N. 17th Street, Suite 1847

Rosslyn, Virginia 22209 USA

VERSION: Final Text (Draft) 5 Sep 2002

## CONTENTS

## FOREWORD

The American College of Radiology (ACR) and the National Electrical Manufacturers Association (NEMA) formed a joint committee to develop a standard for Digital Imaging and Communications in Medicine (DICOM). This DICOM Standard and the corresponding Supplements to the DICOM Standard were developed according to the NEMA procedures.

This Supplement to the Standard is developed in liaison with other standardization organizations including CEN TC251 in Europe and JIRA in Japan, with review also by other organizations including IEEE, HL7 and ANSI in the USA. This Supplement has been prepared by the DICOM Working Group 14 (Security).

The DICOM Standard is structured as a multi-part document using the guidelines established in the following document:

- ISO/IEC Directives, 1989 Part 3 : Drafting and Presentation of International Standards.

This document is a Supplement to the DICOM Standard. It is an extension to PS 3.3, 3.6 and 3.15 of the published DICOM Standard which consists of the following parts:

| | |
|---|---|
| PS 3.1 - | Introduction and Overview |
| PS 3.2 - | Conformance |
| PS 3.3 - | Information Object Definitions |
| PS 3.4 - | Service Class Specifications |
| PS 3.5 - | Data Structures and Encoding |
| PS 3.6 - | Data Dictionary |
| PS 3.7 - | Message Exchange |
| PS 3.8 - | Network Communication Support for Message Exchange |
| PS 3.9 - | Point-to-Point Communication Support for Message Exchange |
| PS 3.10 | Media Storage and File Format for Data Interchange |
| PS 3.11 | Media Storage Application Profiles |
| PS 3.12 | Media Formats and Physical Media for Data Interchange |
| PS 3.13 | Print Management Point-to-Point Communication Support |
| PS 3.14 | Grayscale Standard Display Function |
| PS 3.15 | Security Profiles |
| PS 3.16 | Content Mapping Resource |

These parts are related but independent documents.

## SCOPE AND FIELD OF APPLICATION

The DICOM security extensions proposed in Supplement 31 and 51 enable data confidentiality during DICOM network communication and DICOM media storage interchange, respectively. In both cases, confidentiality (e.g. encryption) always affects complete DICOM SOP instances or network associations – it is not possible to selectively protect only parts of a DICOM SOP instance. In addition, neither of the two extensions can be used if interoperability with existing “non security aware” legacy applications is required.

This supplement adds a mechanism for a selective protection of individual Attributes within arbitrary DICOM SOP Instances. It may be used to achieve protection of identifying information, e.g. a reversible anonymization or pseudonymization of DICOM SOP instances, by implementing modest software changes at the application level only, while continuing to use unmodified lower level message and protocol services for network transfer, storage, and media exchange of composite image information objects.

In particular, protected SOP instances can still be communicated and processed (e.g. displayed) with existing DICOM implementations, security aware or not. Implementations not aware of the security extensions proposed in this supplement will only "see” the anonymized version of the SOP instance, whereas implementations of this supplement may offer functions to reverse/remove the protection if the user has access to the appropriate keys. Use cases in which a selective protection of individual Attributes within DICOM SOP instances may be desirable include:

- Clinical trials, where an authorized entity can view the original data, while other participants in the trial would still be "blind" to the ID of the patient.
- Image databases for case based reasoning, teaching cases or epidemiologic registry.
- Image transmission over high speed networks where the performance penalty for the encryption/decryption of image pixel data may not be acceptable.
- Image interchange scenarios in which the protected images must be processed (transmitted, viewed) with legacy applications and where a selective protection of SOP Instances is considered appropriate, e.g. Teleradiology second opinion requests for modalities where the identification of the patient from the image data itself is highly unlikely.
- Remote service applications, where objects need to be made available to service personel that have no need of the identification information

The underlying principle of the security extension proposed in this supplement is that all DICOM Attribute values to be protected are removed from the SOP instance and, if the Attribute is required for the IOD, replaced by a pseudonym (“dummy value”). The original Attribute values are encrypted and stored in a separate “container” (the Encrypted Attributes Sequence) which is added to the SOP instance. Decryption of the encrypted information requires access to the private “recipient key” which, as with all applications of public key cryptography, is never transmitted.

This supplement addresses the issue of data confidentiality (as defined in ISO 7498-2) at the application level by defining:

- a framework for the protection of Attributes within a Data Set,
- a means of communicating the encryption keys to the intended recipients by means of key transport, key agreement or symmetric key-encryption key schemes,

- an industry standard symmetric encryption algorithm for protection of encrypted Attributes (Triple-DES);
- a baseline profile of standard Attributes that if present, need to be encrypted to provide confidentiality.

Other aspects of security are not addressed by this supplement. In particular, the issues of security policy, guidelines, and the public key infrastructure that is a prerequisite to the use of these security extensions are considered to be beyond the scope of the DICOM standard.

The Profile introduced in this Supplement covers only the general case of creating an alternate representation of an already existing SOP Instance. The Profile does not have a sufficient set of rules that would allow the de-identified SOP Instance to act as the primary representation of the data. It assumes that the original SOP Instance continues to serve as the primary clinical data, and that the de-identified version is for target, special use cases such as a teaching file, for maintenance support, for use in research or clinical trials, etc.

Since this document proposes changes to existing Parts of DICOM the reader should have a working understanding of the Standard. This proposed Supplement includes a number of Addenda to existing Parts of DICOM :

- PS 3.3 Addendum: Attribute Level Confidentiality Extensions to SOP Common Module

- PS 3.6 Addendum: Attribute Level Confidentiality Data Dictionary

- PS 3.15 Addendum: Attribute Level Confidentiality Security Profiles

# 2 Additions to PS 3.3

*Item 2.1 Add the following references to Section 2*

ISO 7498-2 Information processing systems - Open Systems Interconnection - Basic reference Model - Part 2: Security Architecture

ISO 8825-1 Information technology – ASN.1 encoding rules: Specification of Basic Encoding Rules (BER), Canonical Encoding Rules (CER) and Distinguished Encoding Rules (DER).

RFC-2313 PKCS #1: RSA Encryption, Version 1.5, March 1998.

RFC-2630 Cryptographic Message Syntax, June 1999

ANSI X9.52 American National Standards Institute. ANSI X9.52-1998, Triple Data Encryption Algorithm Modes of Operation. 1998.

*Item 2.2 Add to following subsection to Section 3*

## 3.1 REFERENCE MODEL DEFINITIONS

This Part of the Standard makes use of the following terms defined in ISO 7498-2:

a. Data Confidentiality

Note: The definition is “the property that information is not made available or disclosed to unauthorized individuals, entities or processes.”

b. Data Origin Authentication

Note: The definition is “the corroboration that the source of data received is as claimed.”

c. Data Integrity

Note: The definition is “the property that data has not been altered or destroyed in an unauthorized manner.”

d. Key Management

Note: The definition is “the generation, storage, distribution, deletion, archiving and application of keys in accordance with a security policy.”

*Item 2.3 Add the following rows to Table C.12-1 SOP Common Module Attributes*

| **Attribute Name** | Tag | Type | Attribute Description |
|---|---|---|---|
| Encrypted Attributes Sequence | (0400,0500) | 1C | Sequence of Items containing encrypted DICOM data. One or more Items shall be present. Required if application level confidentiality is needed and certain recipients are allowed to decrypt all or portions of the Encrypted Attributes Data Set. See C.12.1.1.x.1 |
| >Encrypted Content Transfer Syntax UID | (0400,0510) | 1 | Transfer Syntax used to encode the encrypted content. Only Transfer Syntaxes that explicitly include the VR and use Little Endian encoding shall be used. |
| >Encrypted Content | (0400,0520) | 1 | Encrypted data. See C.12.1.1.x.2 |

*Item 2.4 Add the following new section C.12.1.1.x*

### C.12.1.1.x Encrypted Attribute Descriptions

#### C.12.1.1.x.1 Encrypted Attributes Sequence

Each Item of the Encrypted Attributes Sequence (0400,0500) contains an encrypted DICOM dataset containing a single instance of the Encrypted Attributes Data Set (Table C.12-X). It also contains encrypted content-encryption keys for one or more recipients. The encoding is based on the Enveloped-data Content Type of the *Cryptographic Message Syntax* defined in RFC 2630. It allows to encrypt the embedded Data Set for an arbitrary number of recipients using any of the three key management techniques supported by RFC 2630:

- *Key Transport:* the content-encryption key is encrypted in the recipient's public key;
- *Key Agreement:* the recipient's public key and the sender's private key are used to generate a pairwise symmetric key, then the content-encryption key is encrypted in the pairwise symmetric key; and
- *Symmetric key-encryption Keys:* the content-encryption key is encrypted in a previously distributed symmetric key-encryption key.

A recipient decodes the embedded Encrypted Attributes Data Set by decrypting one of the encrypted content-encryption keys, decrypting the encrypted dataset with the recovered content-encryption key, and then decoding the DICOM dataset using the Transfer Syntax specified in Encrypted Content Transfer Syntax UID (0400,0510).

Multiple Items may be present in the Encrypted Attributes Sequence. The different Items may contain Encrypted Attributes Data Sets with the same or different sets of Attributes and may contain encrypted content-encryption keys for the same or different sets of recipients. However, if the same Attribute is contained in more than one embedded Encrypted Attributes Data Set, the value of the Attribute must be identical in all embedded Encrypted Attributes Data Sets in which the Attribute is contained.

Note: If the Encrypted Attributes Sequence contains more than one Item, and a recipient holds the key for more than one of the items, the recipient may either decode any single one or more of the embedded

Data Sets at its own discretion. Since the same Attribute is required to have the same value in all embedded Encrypted Attributes Data Sets, it is safe to “overlay” multiple embedded Encrypted Attributes Data Sets in an arbitrary order upon decoding.

### C.12.1.1.x.2 Encrypted Content

The Encrypted Content (0400,0520) Attribute contains an Enveloped-data content type of the cryptographic message syntax defined in RFC 2630. The encrypted content of the Enveloped-data content type is an instance of the Encrypted Attributes Data Set as shown in Table C.12-X (i.e., it is a Sequence with a single Item), encoded with the Transfer Syntax specified by the Encrypted Content Transfer Syntax UID (0400,0510) Attribute. Figure 1 shows an example of how the Encrypted Content is encoded. The exact use of this Data Set is defined in the Attribute Confidentiality Profiles in PS 3.15.

Since the de-identified SOP Instance is a significantly altered version of the original Data Set, it is a new SOP Instance, with a SOP Instance UID that differs from the original Data Set.

Note: 1. Content encryption may require that the content (the DICOM Data Set) be padded to a multiple of some block size. This shall be performed according to the Content-encryption Process defined in RFC-2630.

2. Any standard or private Transfer Syntax may be specified in Encrypted Content Transfer Syntax UID (0400,0510) unless encoding is performed in accordance with an Attribute Confidentiality Profile that specifies additional restrictions. In general, an application entity decoding the Encrypted Attributes Sequence may not assume any particular Transfer Syntax or set of Transfer Syntaxes to be used with Encrypted Content Transfer Syntax UID (0400,0510).

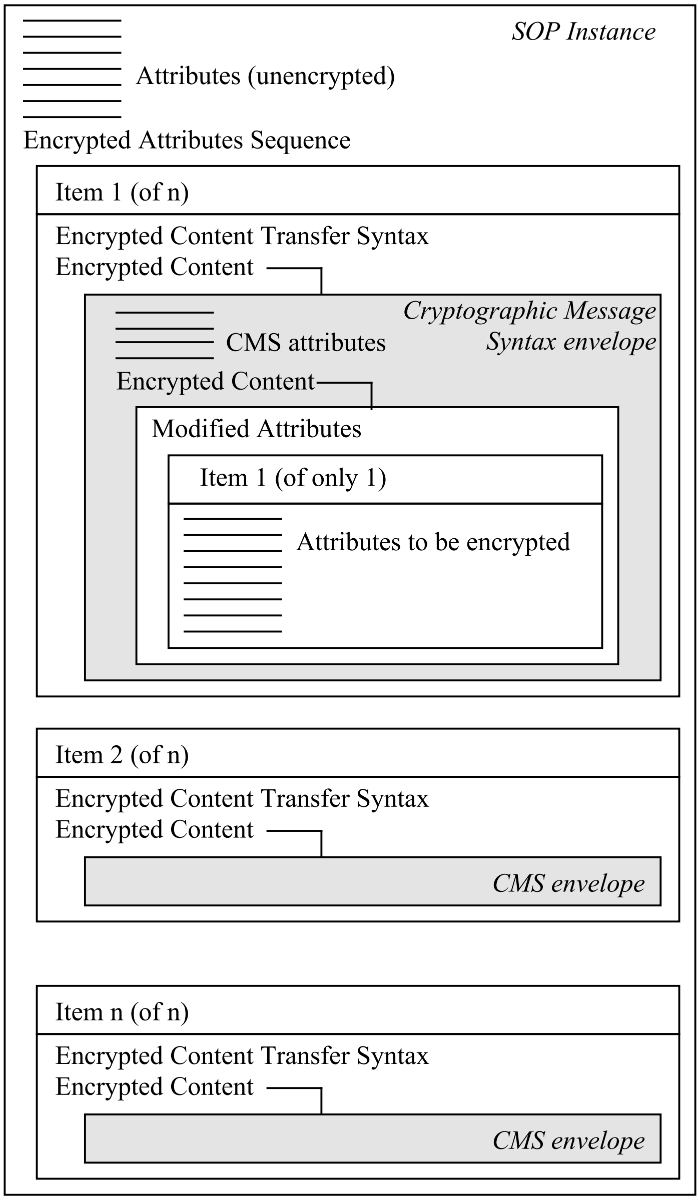
3. For certain applications it might be necessary to “blacken” (remove) identifying information that is burned in to the image pixel data. The Encrypted Attributes Data Set does not specify a means of restoring the original image information without the complete image pixel data being encoded inside the Modified Attributes Sequence (0400,0550). If access to the original, unmodified pixel data is required and the image pixel data cannot be replicated inside the Modified Attributes Sequence (0400,0550) due to resource considerations, the SOP Instance UID may be used to locate the original SOP Instance from which the de-identified version was derived.

4. There is no guarantee that the original SOP Instance can be reconstructed from the data in Encrypted Content. If access to the original data is required, the (de-encrypted) UIDs may be used to locate the original SOP Instance from which the de-identified version was derived.

**Table C.12-X**
**ENCRYPTED ATTRIBUTES DATA SET ATTRIBUTES**

| Attribute Name | Tag | Type | Attribute Description |
|---|---|---|---|
| Modified Attributes Sequence | (0400,0550) | 1 | Sequence of Items containing all Attributes that were removed or replaced by “dummy values” in the main dataset during de-identification of the SOP instance. Upon reversal of the de-identification process, the Attributes are copied back into the main dataset, replacing any dummy values that might have been created. Only a single Item shall be present. |
| *> Any Attribute from the main dataset that was modified or removed during the de-identification process.* | | *3* | |

**Figure 1: Example encoding of Encrypted Attributes Data Set (Informative)**



## 3 Additions to PS 3.6

*Item 3.1 Add the following rows to the table in Section 6*

| Tag | Name | VR | VM |
|---|---|---|---|
| (0400,0500) | Encrypted Attributes Sequence | SQ | 1 |
| (0400,0510) | Encrypted Content Transfer Syntax UID | UI | 1 |
| (0400,0520) | Encrypted Content | OB | 1 |
| (0400,0550) | Modified Attributes Sequence | SQ | 1 |
|  |  |  |  |

## 4 Additions to PS 3.15

*Item 4.1 Add the following new Annex to PS 3.15 Security Profiles (possibly in Annex A Secure Use Profiles)*

## ANNEX X - ATTRIBUTE CONFIDENTIALITY PROFILES

### X.1 BASIC APPLICATION LEVEL CONFIDENTIALITY PROFILE

This Basic Application Level Confidentiality Profile addresses the following aspects of security:

— Data Confidentiality at the application level.

Other aspects of security not addressed by this profile, that may be addressed elsewhere in the standard include:

— Confidentiality in other layers of the DICOM model;

— Data Integrity.

This Profile is targeted toward creating a special purpose, de-identified version of an already-existing Data Set. It is not intended to replace the original SOP Instance from which the de-identified SOP Instance is created, nor is it intended to act as the primary representation of clinical Data Sets in image archives. The de-identified SOP Instances are useful, for example, in creating teaching or research files, where the identity of the patient should be protected, but still be accessible to authorized personnel.

#### X.1.1 De-Identifier

An Application may claim conformance to the Basic Application Level Confidentiality Profile as a de-identifier if it protects *all* Attributes that might be used by unauthorized entities to identify the patient. Protection in this context is defined as the following process:

1. The application may create one or more instances of the Encrypted Attributes Data Set and copy Attributes to be protected into the (single) item of the Modified Attributes Sequence (0400,0550) of one or more of the Encrypted Attributes Data Set instances.

   Note: A complete reconstruction of the original Data Set may not be possible; however, Attributes (e.g. SOP Instance UID) in the Modified Attributes Sequence of an Encrypted Attributes Data Set may refer back to the original SOP Instance holding the original Data Set.

2. Each Attribute to be protected shall then either be removed from the dataset, or have its value replaced by a different “replacement value” which does not allow identification of the patient.

   Note: 1. It is the responsibility of the de-identifier to ensure that this process does not negatively affect the integrity of the Information Object Definition, i.□e. Dummy values may be necessary for Type 1 Attributes that are protected but may not be sent with zero length, and are to be stored or exchanged in encrypted form by applications that may not be aware of the security machanism.

   2. The standard does not mandate the use of any particular dummy value, and indeed it may have some meaning, for example in a data set that may be used for teaching purposes, where the real patient identifying information is encrypted for later retrieval, but a meaningful alternative form of identification is provided. For example, a dummy Patient’s Name (0010,0010) may convey the type of pathology in a

teaching case. It is the responsibility of the de-identifier to ensure that the dummy values cannot be used to identify the patient.

3. It is the responsibility of the de-identifier to ensure the consistency of dummy values for Attributes such as Study Instance UID (0020,000D) or Frame of Reference UID (0020,0052) if multiple related SOP Instances are protected.

4. This standard does not allow selective protection of parts of a Sequence of Items. If an Attribute to be protected is contained in a Sequence of Items, the complete Sequence of Items needs to be protected.

5. The de-identifier should ensure that identifying information that is burned in to the image pixel data is “blackened” (removed). The means by which identifying information is located and removed is outside the scope of this standard.

3. At the discretion of the de-identifier, Attributes may be added to the dataset to be protected.

Note: As an example, the Attribute Patient’s Age (0010,1010) might be introduced as a replacement for Patient’s Birth Date (0010,0030) if the patient’s age is of importance.

4. All instances of the Encrypted Attributes Data Set shall be encoded with a DICOM Transfer Syntax, encrypted , and stored in the dataset to be protected as an Item of the Encrypted Attributes Sequence (0400,0500). The encryption shall be done using RSA [RFC 2313] for the key transport of Triple-DES content-encryption keys. The Triple-DES key length is 168 bits as defined by ANSI X9.52. Encoding shall be performed according to the specifications for RSA Key Transport in RFC-2630.

Note: 1. Each item of the Encrypted Attributes Sequence (0400,0500) consists of two Attributes, Encrypted Content Transfer Syntax UID (0400,0510) containing the UID of the Transfer Syntax that was used to encode the instance of the Encrypted Attributes Data Set, and Encrypted Content (0400,0520) containing the block of data resulting from the encryption of the Encrypted Attributes Data Set instance.

2. RSA key transport of Triple-DES content-encryption keys is specified as a requirement in the European Prestandard ENV 13608-2: Health Informatics - Security for healthcare communication – Part2: Secure data objects.

3. No requirements on the size of the asymmetric key pairs used for RSA key transport are defined in this confidentiality scheme.Implementations claiming conformance to the Basic Application Level Confidentiality Profile as a de-identifier shall always protect (e.g. encrypt and replace) the SOP Instance UID (0008,0018) Attribute as well as all references to other SOP Instances, whether contained in the main dataset or embedded in an Item of a Sequence of Items, that could potentially be used by unauthorized entities to identify the patient.

Note: In the case of a SOP Instance UID embedded in an item of a sequence, this means that the enclosing Attribute in the top-level data set must be encrypted in its entirety.

The Attributes listed in Table X.1-1 contained in Standard IODs typically need to be protected to provide a minimal level of confidentiality from identification. An implementation claiming conformance to the Basic Application Level Confidentiality Profile as a de-identifier shall protect all instances of the Attributes listed in Table X.1-1, whether contained in the main dataset or embedded in an Item of a Sequence of Items, unless the implementation can ensure that the content of these Attributes cannot be used by unauthorized entities to identify the patient.

Notes: 1. The Attributes listed in Table X.1-1 may not be sufficient to guarantee confidentiality of patient identity. In particular, identifying information may be contained in Private Attributes, Curves or Overlays. It is the responsibility of the de-identifier to ensure that all identifying information is removed.

2. It should be noted that conformance to the Basic Application Level Confidentiality Profile does not necessarily guarantee confidentiality. Any encryption scheme may be vulnerable to attack. Also, an organization's Security Policy and Key Management policy are recognized to have a much greater impact on the effectiveness of protection.

3. If the image pixel data contains 'burned in' identifications, the de-identifier may 'black' them out to de-identify the pixel data.

4. National and local regulations, which may vary, might require that additional attributes be de-identified.

**Table X.1-1**
**Basic Application Level Confidentiality Profile Attributes**

| **Attribute Name** | **Tag** |
|---|---|
| Instance Creator UID | (0008,0014) |
| SOP Instance UID | (0008,0018) |
| Accession Number | (0008,0050) |
| Institution Name | (0008,0080) |
| Institution Address | (0008,0081) |
| Referring Physician's Name | (0008,0090) |
| Referring Physician's Address | (0008,0092) |
| Referring Physician's Telephone Numbers | (0008,0094) |
| Station Name | (0008,1010) |
| Study Description | (0008,1030) |
| Series Description | (0008,103E) |
| Institutional Department Name | (0008,1040) |
| Physician(s) of Record | (0008,1048) |
| Performing Physicians' Name | (0008,1050) |
| Name of Physician(s) Reading Study | (0008,1060) |
| Operators' Name | (0008,1070) |
| Admitting Diagnoses Description | (0008,1080) |
| Referenced SOP Instance UID | (0008,1155) |
| Derivation Description | (0008,2111) |
| Patient's Name | (0010,0010) |
| Patient ID | (0010,0020) |
| Patient's Birth Date | (0010,0030) |
| Patient's Birth Time | (0010,0032) |
| Patient's Sex | (0010,0040) |
| Other Patient Ids | (0010,1000) |
| Other Patient Names | (0010,1001) |
| Patient's Age | (0010,1010) |

| | |
|---|---|
| Patient's Size | (0010,1020) |
| Patient's Weight | (0010,1030) |
| Medical Record Locator | (0010,1090) |
| Ethnic Group | (0010,2160) |
| Occupation | (0010,2180) |
| Additional Patient's History | (0010,21B0) |
| Patient Comments | (0010,4000) |
| Device Serial Number | (0018,1000) |
| Protocol Name | (0018,1030) |
| Study Instance UID | (0020,000D) |
| Series Instance UID | (0020,000E) |
| Study ID | (0020,0010) |
| Frame of Reference UID | (0020,0052) |
| Synchronization Frame of Reference UID | (0020,0200) |
| Image Comments | (0020,4000) |
| Request Attributes Sequence | (0040,0275) |
| UID | (0040,A124) |
| Content Sequence | (0040,A730) |
| Storage Media File-set UID | (0088,0140) |
| Referenced Frame of Reference UID | (3006,0024) |
| Related Frame of Reference UID | (3006,00C2) |

### X.1.2 Re-Identifier

An Application may claim conformance to the Basic Application Level Confidentiality Profile as a re-identifier if it is capable of removing the protection from a protected SOP instance given that the recipient keys required for the decryption of one or more of the Encrypted Content (0400,0520) Attributes within the Encrypted Attributes Sequence (0400,0500) of the SOP instance are available. Removal of protection in this context is defined as the following process:

1. The application shall decrypt, using its recipient key, one instance of the Encrypted Content (0400,0520) Attribute within the Encrypted Attributes Sequence (0400,0500) and decode the resulting block of bytes into a DICOM dataset using the Transfer Syntax specified in the Encrypted Content Transfer Syntax UID (0400,0510).

   Note: If the application is able to decode more than one instance of the Encrypted Content (0400,0520) Attribute within the Encrypted Attributes Sequence (0400,0500), it is at the discretion of the application to choose any one of them.

2. The application shall move all Attributes contained in the single item of the Modified Attributes Sequence (0400,0550) of the decoded dataset into the main dataset, replacing"dummy value" Attributes that may be present in the main dataset.

   Notes: 1. Re-identification does not imply a complete reconstruction of the original SOP Instance, since it is not required that all Attributes being protected be part of the Encrypted Attributes Data Set. If the original UIDs are part of the Encrypted Attributes Data Set, they might be usable to gain access to the original, unprotected SOP Instance.

2. The presence of an encrypted data set that cannot be decrypted indicates that some or all of the attribute values in the message may not be real (they are dummies). Therefore, the recipient must not assume that any value in the message is diagnostically relevant.

### X.1.3 Conformance Requirements

The Conformance Statement of an application that claims conformance to the Basic Application Level Confidentiality Profile shall describe:

- which Attributes are removed during protection;
- which Attributes are replaced by dummy values and how the dummy values are generated;
- which Attributes are included in Encrypted Attributes Data Sets for later re-identification, and any pertinent details about how keys are selected for performing the encryption;
- whether or not the application is able to ensure integrity of dummy values for references such as SOP Instance UID, Frame of Reference UID, etc. if multiple SOP instances are protected;
- which Attributes and Attribute values are inserted during protection of a SOP instance;
- which Transfer Syntaxes are supported for encoding/decoding of the Encrypted Attributes Data Set;
- which Confidentiality Schemes are supported;
- any additional restrictions (e.□g. key sizes for public keys).

## 5 Index of Attribute Tags