Integrating the Healthcare Enterprise

IHE Quality, Research and Public Health (QRPH)

White Paper

Using IHE profiles for Healthcare - Secondary Data Access

Published

Revision 1.1

Date: November 9, 2016

Author: QRPH Technical Committee

Email: qrph@ihe.net

Please verify you have the most recent version of this document. See here for Published versions and here for Public Comment versions.

Foreword

Integrating the Healthcare Enterprise (IHE) is an international initiative to promote the use of standards to achieve interoperability among health information technology (HIT) systems and effective use of electronic health records (EHRs). IHE provides a forum for care providers, HIT experts and other stakeholders in several clinical and operational domains to reach consensus on standards-based solutions to critical interoperability issues.

The primary output of IHE is system implementation guides, called IHE Profiles. IHE publishes each profile through a well-defined process of public review and trial implementation and gathers profiles that have reached final text status into an IHE Technical Frameworks.

This white paper is published on November 9, 2016. Comments can be submitted at http://www.ihe.net/QRPH_Public_Comments.

General information about IHE can be found at: http://ihe.net.

Information about the IHE IT Infrastructure domain can be found at: http://ihe.net/IHE_Domains.

Information about the organization of IHE Technical Frameworks and Supplements and the process used to create them can be found at: http://ihe.net/IHE_Process and http://ihe.net/Profiles.

The current version of the IHE IT Infrastructure Technical Framework can be found at: http://ihe.net/Technical_Frameworks.

CONTENTS

1 Introduction

This document, "Using IHE profiles for Healthcare-Secondary Data Access" White Paper, describes how to build a standardized infrastructure using IHE profiles to allow a Secondary Data Usage Community to have access to the information (both data and documents) available in a Clinical Community and vice versa.

With the term "Secondary Data Usage Community", we mean a Research Organization, a Public Health Organization, an Epidemiology Organization, a Quality Reporting Agency or any other organization which collects and uses a patient's data for purposes different to the direct care of the patient and where personal data identifying specific patients might not be allowed and/or necessary: for example, research, population health management, health service management and quality assurance, public health surveillance, disease control, public safety emergency, education, market studies, and enabling the payment of care provision[1].

With the term "Clinical Community", we mean any group of healthcare enterprises, which work together for the patient's direct care and share common policies and a common infrastructure, for example, Health Information Exchange Systems (HIE Systems).

A Secondary Data Usage Community may be interested to have access to clinical data available in a Clinical Community, for example, to perform research studies on real world data, but even a Clinical Community may be interested in data owned by the Secondary Data Usage Community (as Case Report Forms or Health Quality Reports). In this white paper, it will be described how to allow the data access from another community in a standard, reliable and secure way in respect of the patient's privacy: access to data is available even if a community is not allowed to know the patient's identifier used by the other community.

1.1 Expected Knowledge and References

It is assumed that the reader has a working knowledge of the following integration profiles defined within the IT Infrastructure Technical Framework available at http://www.ihe.net/Technical_Frameworks/#IT.

- Cross-Enterprise Document Sharing (XDS)
- Patient Identifier Cross-Referencing (PIX)
- Cross-Community Access (XCA)
- Patient Demographic Query (PDQ)
- Cross-Community Document Reliable Interchange (XCDR)

[1] In this white paper, purpose is defined according to the standard DD ISO/TS 14265:2011 "Health informatics — Classification of purposes for processing personal health information"

The reader is also referred to the following supplements to the Trial Implementation QRPH Technical Framework available at http://www.ihe.net/Technical_Frameworks/#qrph:

- Redaction Services (RSP)
- Data Element Exchange (DEX)
- Clinical Research Process Content (CRPC)
- Retrieve Protocol for Execution (RPE)
- Aggregate Data Exchange (ADX)

Finally, the reader is referred to the Query for Existing Data (QED) supplement to the IHE PCC Technical Framework (http://www.ihe.net/Technical_Frameworks/#pcc).

1.2 Purpose of the "Using IHE profiles for Healthcare-Secondary Data Access" White Paper

This document, "Using IHE profiles for Healthcare-Secondary Data Access" White Paper, describes how to build a standardized infrastructure using IHE profiles to allow a Secondary Data Usage Community to have access to the information (both documents and data within documents) available in a Clinical Community and vice versa. The solutions herein described allow managing data access in respect of privacy issues and rules agreed by the two communities. The Secondary Data Usage Community is allowed to have access to clinical data available in the Clinical Community even in the case that the patient's identity shall not be disclosed: in this case, the clinical documents shall be de-identified before being provided to the Secondary Data Usage Community. In the same way, the solutions herein presented allow a Clinical Community to have access to documents stored in the Secondary Data Usage Community even if the Clinical Community is not allowed to know the patient identifier used by the Secondary Data Usage Community.

This white paper addresses, in particular, the need and great interest shown in the last years by Clinical Research Organizations, Public Health institutions, Epidemiology organizations to have access to the huge and rich amount of information available in Clinical Communities in a fast, reliable, and secure way. In the last years, lots of Health Information Exchange (HIE) Systems have been established in Clinical Communities: they store a huge amount of clinical data with a patient-centric vision and with a high level of organization.

A scenario of interest is depicted in Figure 1.2-1 and shows two communities: a HIE system (Clinical Community) and a research organization (Secondary Data Usage Community). The two communities store data related to the same people but identified with different patient identifiers for privacy reasons. Both of the communities would like to have access to data available in the other community. In this example a researcher would like to have access to clinical data available in the HIE system to calculate the prevalence of a disease and a family doctor would like to have access to Case Report Forms collected during clinical trials related to the patient he is assisting: this information may allow him to make a prompt diagnosis and avoid useless

exams. This white paper describes how the two communities can have access to data available in the other community.

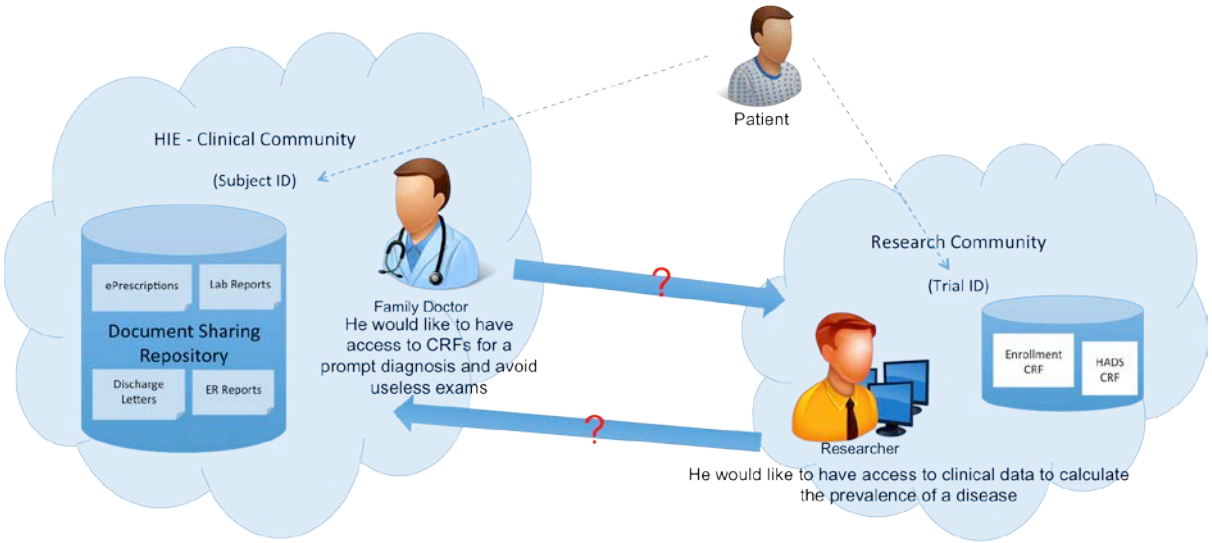


Figure 1.2-1: Scenario of Interest

This white paper does not define any requirement about the type of infrastructure characterizing the communities involved: they can be either Document Sharing environments (e.g., a Cross-Enterprise Document Sharing – XDS- environment) or not. This document shows how communication between communities can be performed even if they are not Document Sharing environments: the communication can be managed through the application of the Cross-Community Access (XCA) Profile, as further discussed later throughout this document.

This document provides guidance about general and fundamental privacy issues (in particular data de-identification for secondary data usages) related to the different use cases and it shows how to manage them from a technical point of view. The Section 2.4 and Appendix A discuss general privacy issues related to secondary data usage according to reference standard-regulations and main international legislations. This document is not focused on other specific privacy and security requirements (e.g., patient's consent management and access control systems and policies), which rely on local policies and infrastructures, however Appendix B provides some guidance about how these requirements can be implemented in a standard way compliant to IHE profiles.

This white paper focuses on how a community can have access to documents (or to single data stored within documents) available in another community. It is out of the scope to describe the process of document creation even if it involves the two communities. For example, it is out of the scope of this white paper to describe how a Case Report Form designed by a Research Organization is completed, even if it is populated through an Electronic Data Capture system available in a EHR within a Clinical Community.

This white paper focuses on access to patient-level documents and to single specific data stored within structured documents. Since data access requires first of all document access and then data extraction, herein, if not otherwise specified, when "document access" is indicated, also "data access" is meant. However, it is out of scope the description about how to access documents related to a group of patients (population-level documents) or to documents not related to any specific patient/population (e.g., clinical guidelines documents). Even if these kinds of documents are available in "Secondary Data Usage Communities", until now no IHE profiles exist to allow querying and retrieving of these kinds of documents.

The description of interactions between two Clinical Communities is out-of-scope for this white paper, since this subject is addressed by other IHE documents, as the ITI XCA Profile and the ITI "Cross-Community Information Exchange" white paper.

1.3 Use-cases

1.3.1 Use case 1: Epidemiologycal study

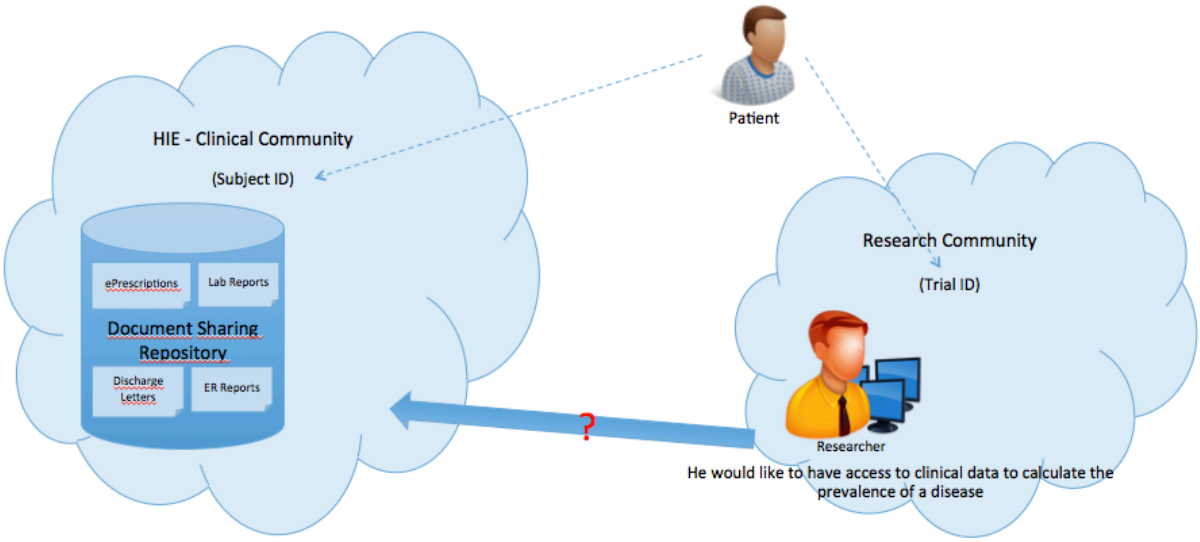


Figure 1.3.1-1: Epidemiologycal Study (Use Case 1)

A researcher would like to calculate the prevalence of a disease within his/her community and, in the current state, it is not a trivial task since an observational study is usually performed and cumbersome procedures have to be established: they take a long time, include a long follow-up, and involve a lot of people and resources. For example, in order to evaluate the prevalence of gestational diabetes mellitus in relation to race and socioeconomic status in the region where he lives, a cross-sectional study is usually performed: lots of women are enrolled in the study, many

hospitals may be involved in order to reach a sufficient sample size and women should be screened for all the pregnancy period. In the HIE system established in the Region different kinds of documents are produced: Discharge Summaries, ER Referrals, ePrescriptions, eReferrals, Laboratory Reports, Pathological Anatomy Reports, Vaccination reports, which may contain the diagnosis information, which would allow the researcher to calculate the prevalence of a specific diagnosis. However, in the current state the researcher is not allowed to retrieve data from the HIE system. This use-case is depicted in Figure 1.3.1-1.

In the desired future state, the researcher asks the administrative authority of the HIE system to have access (in respect of legal and privacy issues) to the clinical data needed to answer the research question. Once the permission is obtained and an agreement is established by the two communities, a standard infrastructure can be implemented according to the solution presented in this white paper, which allows the researcher to get data stored in the HIE system in a secure way, particularly in respect of the patient's privacy. He/she can ask either for patient-level data and then perform the analysis by him/herself to get the final aggregate results or directly for aggregate data.

1.3.2 Use case 2: CRFs retrieval for clinical purposes

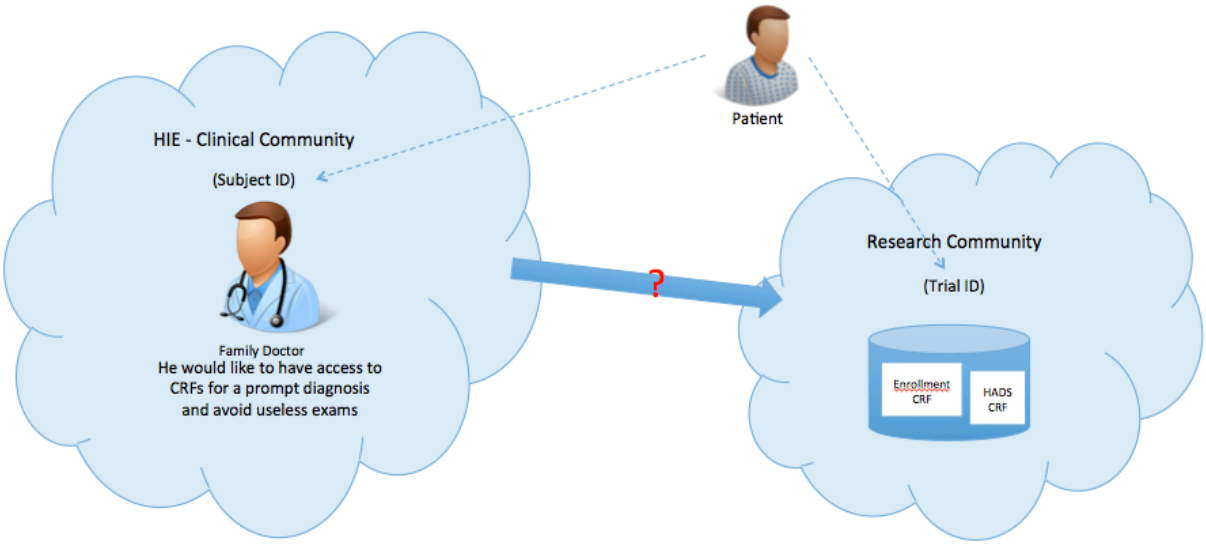


Figure 1.3.2-1: CRFs Retrieval For Clinical Purposes (Use Case 2)

A clinician working in the clinical community would like to have access to data collected during clinical trials and stored in the research community. During clinical trials lots of Case Report Forms (CRFs) have been administered to patients enrolled in the study, containing, for example, their clinical parameters values, main clinical events related to the study outcome, quality of life

level, behavioral habits, psychological and social information. The following example offers a good illustration of this use-case: a patient enrolled in a clinical trial (performed by the research organization) goes to his family doctor because of a thoracic pain. The patient tells the doctor about his participation in a clinical trial about a new drug meant to reduce anxiety. The doctor, who does not understand the cause of the pain, would like to have access to the patient's data collected during the trial, especially to CRFs with anxiety information not available in the patient's EHR. However, in the current state the doctor is not allowed to retrieve data stored in the research community. This use-case is depicted in Figure 1.3.2-1.

In the desired use case, after an agreement is established between the communities to allow the HIE system to have access to clinical data stored in the Research Organization, the doctor can retrieve data stored amongst in the research organization. This white paper provides a standard solution to allow the mechanism of access by the doctor to the patient's CRFs stored in the research organization: in particular, since the doctor is not allowed to know the patient's Study ID used by the Research Organization during the clinical trial, data are provided replacing the Study ID with the patient's identifier known by the HIE system.

1.4 Intended Audience

The intended audience of the "Using IHE profiles for Healthcare-Secondary Data Access" White Paper is:

- Public Health Institutions who want to integrate National/Regional Health Information Exchange systems with Secondary Data Usage Communities to allow the re-use of clinical data for research/quality/epidemiology purposes;
- Managing/IT staff of healthcare institutions who want to integrate with other institutions performing research/quality/epidemiology activities;
- Research Organizations/Quality Agencies/Epidemiology Institutions who want to integrate with healthcare institutions to perform research activities on real-world data, healthcare quality assessment, epidemiology studies, etc.
- Experts involved in standards development.

1.5 Comment Process

IHE International welcomes comments on this document and the IHE initiative. They can be submitted by sending an email to the co-chairs and secretary of the Quality, Research and Public Health domain committees at qrph@ihe.net.

1.6 Open and Closed Issues

NA

1.7 Glossary[2]

Anonymity: Anonymity means that the subject is not identifiable. For example, a patient cannot be identified from a teaching file. From the perspective of an attacker, anonymity means that no individual subjects can be identified.

Anonymization: A process that is intended to irreversibly remove the association between a subject and information that can identify the subject. If the process is intended to be reversible and a new identifier is substituted for the subject's real identifiers, then the process is called pseudonymization.

Anonymous identifier: An identifier for a subject that, in contrast to pseudonymization, is not intended to allow relinking to the subject. It may be created from one-way mapping from a subject to an identifier that cannot be reversed. This is different than pseudonymization, see below.

De-identification: Any process that removes the association between a subject's identity and the subject's data elements. Anonymization and pseudonymization are types of de-identification.

The major algorithms used in de-identification are:

- Fuzzing – Adding "noise" to data
- Redaction – Removing data, or replacing it with missing data indicators
- Generalization – Making data less specific
- Longitudinal consistency - Modifying data so that data from many records remain consistent.
- Text Processing – Manual processing for free-format text
- Pass-through – Unmodified data is preserved in the resulting dataset

Direct identifying data: Data that directly identifies a single individual. Direct identifiers include data that can be cross-referenced through commonly available information sources, e.g., telephone number. Locally used identifiers (such as hospital IDs) can be considered directly identifying to personnel of the local domain.

Explicit consent: express consent permission that is freely and directly given, expressed either orally or writing.

Identifiable person: A person who can be identified, directly or indirectly. For example through one or more factors specific to their physical, physiological, mental, economic, cultural or social identity (see "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data").

[2] Definitions indicated in IHE IT Infrastructure Handbook – De-Identification

Implied consent: A voluntary agreement with what is being done or proposed that can be reasonably determined through the actions or inactions of the data subject.

Informed consent: A consent granted on the basis of knowledge.

Indirect identifying data: "Data that does not directly identify a single individual but may be used in collaboration with other indirect identifiers to identify an individual. ... Examples: Zipcode(sic), Sex, Age, Date-of-Birth, Race." [ISO 25237]

Irreversibility: The inability to determine an original value, or set of values. This is not always a simple binary statement. It is often a measure of difficulty. It is computationally difficult to determine the original values once it has been subjected to a SHA-256 one-way hash with a salt. Some national organizations may have the resources to perform this computation, and changes in computer technology will change the degree of difficulty.

Pseudonym: A computed or assigned value that is substituted for one or more data elements in that subject's record. Alias and nickname are common terms for pseudonym. For example, a pseudonym of "csrk123" could be added to a subject's record, and that subject's first, last, middle, and national ID numbers could be removed. The protection provided by a pseudonym is dependent on the system used to create and protect the relationship between the pseudonym and the person's real identity. Well known aliases are an example of pseudonyms that provide little protection, more people know the alias "Lenin" than his birth name. This differs from anonymization by preserving continuity throughout the resulting data set. In this white paper with the term "pseudonym", we refer in general to the subject identifier used by the community asking for data that is usually different to that used by the community providing data.

Pseudonymization: A particular type of anonymization that removes the association between data and a subject and introduces a new identifier that establishes a bidirectional-mapping between that subject and the new identifier. Pronunciation guide: "soo-DON-imm-iza-tion", rhymes with optimization.

2 Concept of Community

In this chapter the definition and description of the "community concept" is provided.

2.1 Community Definition: Borders and Characteristics

A Community, as conceived in this white paper, can be both a Clinical community for the patient's care and a Secondary Data Usage Community performing any other type of activity. The definition of community here is slightly more restrictive than that usually considered in the ITI/PCC IHE domains.

We define a community as follows:

A community is a group of people/facilities/enterprises that have agreed to work together and may produce and/or consume documents. Each community has agreed on a common set of policies for the sharing of documents within the community via an established mechanism. A community is identifiable by a globally unique id (the homeCommunityId). Such communities may be XDS Affinity Domains, which define document sharing using the XDS Profile, or any other communities, regardless of their internal sharing structure. Membership of a person/facility/enterprise in one community does not preclude it from being a member in another community. Each community identifies a patient with one and only one identifier, which can be different or equal to that used by another community. A patient may belong to one or multiple communities. The number of patients can vary between communities; a community can have no patients (no patients' identifiers management) and in this case it can have or be interested to have "cases" (for anonymous data related to unspecified patients). Each community produces, collects and uses documents for one and only one specific purpose.

2.1.1 XDS Compliance

This white paper does not make any requirement about the type of infrastructure characterizing the communities involved: they can be either compliant to the IHE Cross-enterprise Document Sharing (XDS) Profile or not.

Though compliance to the XDS standard is not required, an XDS Environment provides great advantages in terms of high organization and standardization of document management and processes. This white paper shows how the XDS logic can be used for the communities' communication and data access, even if the communities are not established as XDS Environments: In this case, the interface they expose to the exterior world (through gateways) hides the logic used within the communities and translate it to an XDS logic. The XCA standard allows handling this kind of situation, as described later in 3.1.2.1.

2.1.2 Relationship between Communities and Patient Identifiers

As indicated above in the definition of community, each Community identifies a patient with only a Cross-Enterprise subject identifier. This statement means that if an institution performs different activities, each one using a specific subject identifier for a patient and it is not allowed

to link the different subject identifiers related to the same patient (for example in the case of different clinical trials performed by the same Clinical Research Organization), the institution has to be considered composed of many communities, one for each activity. This distinction allows to respect the privacy of patients involved in multiple activities: for example, in the case of clinical trials, if a patient is involved in many studies within the same institution, the institution should not know this information because each study involves anonymous patient identifiers.

If multiple patient identifiers are used within an institution, but only one is the Master Patient Index (MPI), logically and also operatively this is the reference identifier for that subject in that institution: in this case the institution can be still considered a single Community.

The patient identifier used by a Community may be the same of that used by another Community if there are no privacy issues avoiding it: for example, in the case of a Clinical Community using a National Identifier to identify a patient and a Public Health organization which is allowed to receive patient level quality reports (with the patient's National Identifier) from the Clinical Community to evaluate quality indicators about health services provision.

2.1.3 Document Stewardship

In this white paper the document stewardship is related to the place where the document is stored and not, for example, to the place where it has been created. It means that if the process of document creation starts in a Clinical Community (as in the case of a Case Report Form for a clinical trial populated with data already available in the EHR) and then the document is sent to and archived by a Secondary Data Usage Community (the Research Organization performing the clinical trial), the document belongs to the Secondary Data Usage Community. This assumption is related to the purpose of this white paper, which is to allow access to data and documents stored in a Community by another Community and not to allow the complete interoperability between the two communities, including exchange of data for the document creation.

2.2 Clinical and Secondary Data Usage Communities

As stated in the introduction, with the term Clinical Community we mean "any group of healthcare enterprises, which work together for the patient care and share common policies and a common infrastructure".

Examples of Clinical Communities are:

- Regional Health Information Exchange Systems;
- Nationwide Health Information Exchange Systems;
- Specialized or disease-oriented Care (as cardiology specialists and an acute cardiology center, an oncology network, a diabetes network);
- Federation of enterprises (a Regional federation made up of several local hospitals and healthcare providers).

Clinical Communities are patient-centric and create and consume patient-level documents according to rules and procedures established by each community. In the case of Clinical Communities organized as XDS Environments, documents are stored in XDS Document Repositories and indexed by the XDS Document Registry. Documents can have different formats; they can be structured (as CDA®[3] documents) or unstructured documents (as PDF documents).

As previously stated in the introduction, with the term Secondary Data Usage Community we mean "any organization which collects and uses patient's data for purposes different to the direct care of the patient and where identified data might not be allowed and/or necessary: for example research, population health management, health service management and quality assurance, public health surveillance, disease control, public safety emergency, education, market studies, and enabling the payment of care provision".

Examples of Secondary Data Usage Communities are:

- Clinical Research Organizations;
- Public Health Organizations;
- Epidemiology Organizations;
- Quality Reporting Agencies;
- Bio-surveillance Systems;
- Insurance companies.

Secondary Data Usage Communities create and consume documents, which can be related to specific patients in a patient-centric vision (as in Clinical Communities) or to population-level documents or also to any specific patient/population documents (e.g., clinical guidelines). Examples of patient-level documents are Case Report Forms collected during clinical trials, patient-level quality reports (as QRDA category I), and specific documents collected for bio-surveillance purposes (as adverse events reports or reports for healthy weight surveillance). Examples of population-level documents are quality reports with aggregate measures (as QRDA category III), and reports about clinical trial results expressed as aggregate data.

It is out of scope of this white paper to describe the access from an external Community to population-level documents or non-specific patient/population documents stored with the Secondary Data Usage Communities. However this white paper will describe how a Secondary Data Usage Community can ask for data available in clinical patient-level documents stored by a Clinical Community and ask they are provided as aggregated data, so as population-level documents.

[3] CDA is the registered trademark of Health Level Seven International.

Documents stored in the Secondary Data Usage Community can have different formats: structured documents (such as CDA, xml CRFs compliant with CDISC ODM standard, json) or unstructured documents (such as PDF or text documents).

2.3 Number of Communities

This white paper provides guidance about how to build a standard infrastructure supporting the means to query and retrieve data held by another kind of community. However, the number of communities interested to communicate for this purpose can be even larger than two. This is for example the case of a HIE system, which provides real world data both to a public institution performing epidemiology studies and to a private pharmaceutical company performing a clinical trial. According to the definition of community provided above, also when a Clinical Research Organization is performing two different clinical trials and is interested to linkage data collected during the research studies with real world clinical data stored in a HIE systems, three communities instead of two are actually involved. Moreover, if a Clinical Organization performs not only healthcare activities, but also other kinds of activities (as marketing or research), it has to be considered as composed by different communities, one for each kind of activity.

The general situation about interaction between communities is depicted in Figure 2.3-1.

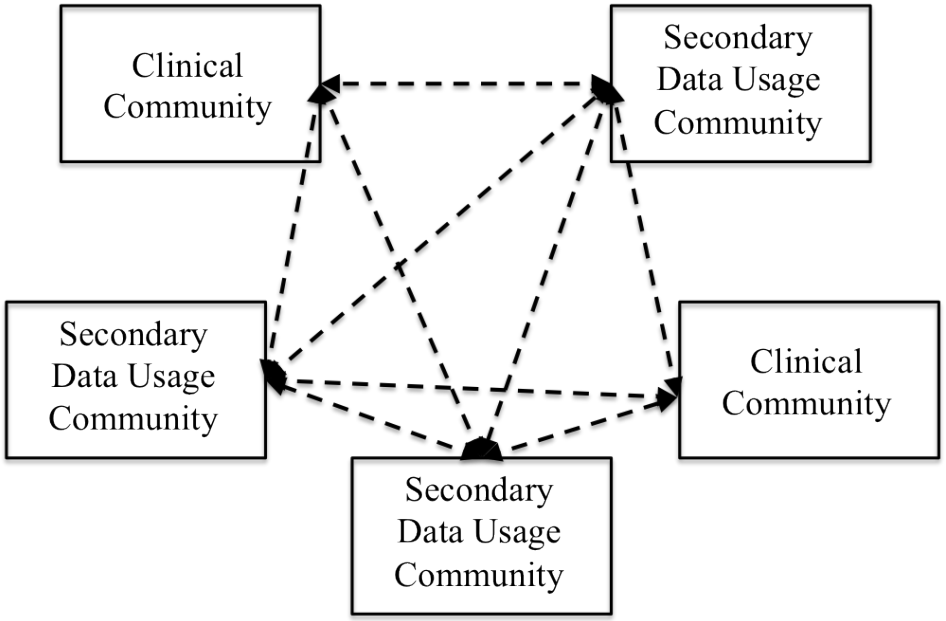


Figure 2.3-1: Interaction between communities

The figure depicts both real scenarios when in a same area different projects regarding cross-community data access coexist and real scenarios when a single project involving data exchange

among multiple communities exists. Interactions between two Clinical Communities are not represented in figure, since this subject is out-of-scope for this white paper.

2.4 Privacy and Security Considerations

A fundamental principle underlying the use of personal health data is:

"It is essential to know the purposes for which data was originally collected and that all subsequent processing activities be the same as, or consistent with, the original purpose… For ethical and legal reasons, it is normally the case that information is used only for the purpose for which it was collected or created. This purpose can be specified explicitly and consented to. Consent to use data for a particular purpose can also be implied, although it is almost always a requirement that the purposes be declared. Where data are intended for further and different purposes, a new purpose can require a new consent. For example, in some jurisdictions, data collected for health care cannot automatically be used for research, nor information collected for research used for care, without obtaining new consent. Knowing the purpose for which access to information is intended is essential in order to determine if access to data for processing activities are appropriate."[4].

Therefore, data may be used for further and different purposes and this is the scenario of interest for this white paper. The list of purposes of health data use, as defined in ISO TS 14265:2011(E), is provided in Table 2.4-1, shown here:

Table 2.4-1: Purpose code and corresponding classification term according to ISO TS 14265:2011(E)Purpose code	Classification term
1	Clinical care provision to an individual subject of care
2	Emergency care provision to an individual subject of care
3	Support of care activities within the provider organization for an individual subject of care
4	Enabling the payment of care provision to an individual subject of care
5	Health service management and quality assurance
6	Education
7	Public health surveillance, disease control
8	Public safety emergency
9	Population health management

[4] ISO/TS 14265:2011: Health informatics — Classification of purposes for processing personal health information

Table 2.4-1: Purpose code and corresponding classification term according to ISO TS 14265:2011(E)Purpose code	Classification term
10	Research
11	Market studies
12	Legal procedure
13	Subject of care uses
14	Unspecified

The use of data for a new purpose can require a new consent. For example, in some jurisdictions, data collected for health care cannot automatically be used for research, nor information collected for research used for care, without obtaining new consent.

Other jurisdictions may not require collecting a new consent as long as data anonimity is guaranteed. ISO TS 14265:2011(E) also states:

"Data purposes or specific uses may or may not require identifiable data. Some data purposes might require the use of identifiable, de-identified, anonymous, pseudonymous or aggregate data, it is commonly understood that where identity is not required it should not be disclosed. Identity is most often required when the purpose of use is to the benefit of the individual data subject, as when the data subject is also a subject of care. The de-identification, anonymization, or pseudonymization of data may be applied as a confidentiality control or condition of use, just as appropriate authority may be applied as a condition of collection, use or disclosure. This in turn means that just as de-identification may be applied as a condition of use, a defined data purpose may be a requirement for the use of even de-identified or anonymized data according to the policy or law of a given jurisdiction."

Therefore, data usually need to be de-identified as anonymous, pseudonymized or aggregate data (depending on jurisdiction and purpose of use) before being re-used. This issue is strongly emphasized in this white paper, which shows standard solutions about how to de-identify data before being provided to another community. The de-identification process in general requires that some data are removed (redaction process), others are elaborated (e.g., generalization algorithms) and others are preserved. Therefore, a "filtering system" between communities is needed to "pass some data", "block others" and "elaborate other" data before providing them to the final community. This white paper describes how to implement this filtering layer in a standard way and also shows how different profiles have to be grouped in order to protect the patient's privacy.

Since the permission to re-use data for a specific purpose does not generally allow the permission of re-use of data for another purpose, a community, as defined in this white paper, is

related to one specific purpose of use. For example, if an organization would like to have access to health data both for research and health quality assurance, it should be clearly and explicitly stated and the organization should be conceived as two different communities with different policies and generally two different consents should be collected according to local jurisdictions.

Knowing the purpose for which access to information is intended is essential in order to determine if access to data for processing activities is appropriate. It is therefore essential to ensure that the context within which access and use is asserted is the correct one. Purpose of use, when clearly defined, helps to ensure that access to protected information items is only to properly authorized users under a specific, appropriate and unambiguous policy.

The process of approval about data/document access for a specific user and the policy management could be performed either by the single communities or by a central system playing the role of a Trusted Third Party. Further details about the access control and policy management are presented in Appendix B.

3 Main Architecture Features

This chapter describes the main features about the standard architecture solution allowing the means to query and retrieve data held by other communities. The main actors involved and their main functionalities are presented and discussed.

3.1 The Trusted Third Party architecture

Since a mechanism of filtering is needed in the interactions between communities in particular to protect the patient's privacy, the architecture solution that will be presented in this white paper is based on the idea of having a central management system. We call the central management system as "Trusted Third Party" (TTP), because all the communities involved trust on this central actor, where all the requests of access to data are sent, analyzed about their adherence to the study definition and, if the request for data is granted, data is retrieved, de-identified and forwarded to the community of interest.

The global architecture described in this white paper is schematically represented in Figure 3.1-1.

The "TTP solution" proposed in this white paper addresses the general requirements of privacy, in particular when many communities are involved. This model is known, for example ISO/TS 25237[5] states "In the case where the pseudonymization service is required to synchronize pseudonyms across multiple entities or enterprises, a trusted service provider may be employed. Trusted services may be implemented through numerous options, including commercial entities, membership organizations, or government entities. Providers of trusted services may be governed through legislation or certification requirements in various jurisdictions." In some particular cases and jurisdictions, for example when only two communities are involved and data shall be provided as anonymous data, one of them can be allowed to perform the anonymization process before delivering data to the other community: in this case the schema represented in Figure 3.1-1 can be simplified and the TTP-as-role will be played by a system in the community providing data. On the contrary, if pseudonymization is needed, a central TTP managing correspondent patient's identifiers in the different communities usually is needed, because a community is not generally allowed to know the patient's identifier used in the other community. The function of the TTP is to retrieve documents, to perform data de-identification and to provide data to the final community; therefore, the storage of data/documents (neither the original documents nor the de-identified documents) shall not be performed by the TTP since the TTP is not allowed to use these data for its own purposes. Some jurisdictions may not allow at all the TTP to have access to clinical data, but only to Personal Identifiable Information (PII) of the patients. This requirement implies that the clinical document has to be split into two parts (one containing PII and one containing clinical data) before being proving to the requesting community. This specific case requires a technical solution different to the general solution

[5] ISO/TS 25237: Health informatics — Pseudonymization

presented in this white paper (in this chapter and in the following ones). Moreover, this specific technical solution needs functionality not already defined by IHE profiles. In Section 3.1.3.2 an overview of a possible solution is presented and the "missing" IHE functionalities are highlighted.

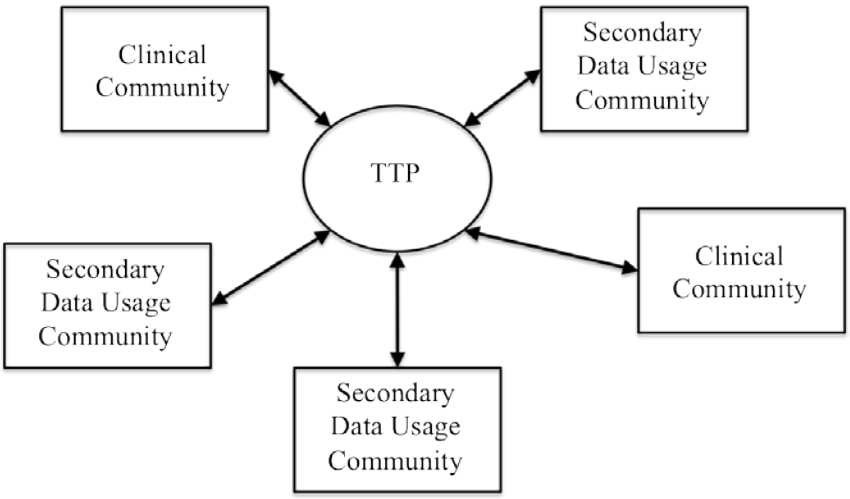


Figure 3.1-1: General architecture

The main functionalities that are performed by the TTP are:

1. **Cross-community study management** (the TTP has to know and store the protocol of the study for which data access from a community to another is needed. The protocol shall include specifically the board approval or, otherwise, a demonstration that the board approval is not needed, eligibility criteria, people involved in the study;
2. **Cross-community data and document access and provision** (the TTP plays the role of a bridge in the communication between the communities and performs the management of the requests/responses for data/documents access. Specifically, it is in charge to receive the requests and forward them to the right responding communities only after the patient identifier and document metadata are translated to those known by the responding community. After it has received data from the responding community, it is in charge to provide them to the requesting community in the desired format and modality. This is a basic functionality regarding the management of transactions and data elaboration, which allows then to perform all the other TTP functionalities listed below);
3. **Cross-community de-identification service** (the TTP is in charge to perform data/document de-identification in order to protect the patients' privacy);

4. **Cross-community patient identity management** (the TTP is in charge to manage the correspondences among all of the identifiers for the same patients in all the communities);
5. **Cross-community semantic service** (the TTP is in charge to know the type and metadata of documents storing data of interest in the different communities).

In the following sections, these functionalities are presented, as well as IHE profiles that can be implemented to address the different needs. From now on, for simplicity, only two communities and the TTP will be considered: the community asking for documents will be called the "Requesting Community" and that providing documents will be called the "Responding Community".

3.1.1 Cross-community Study Management

When a collaboration starts between two or more communities that involves access of data held by a community from another community, first of all a study protocol has to be defined. The main information to be specified in the study protocol is the purpose and outcomes of the study, type of study (e.g., interventional, observational), data of interest, level and type of de-identification to be applied in order to meet ethical and privacy needs, the duration of the study, patients' eligibility criteria, staff and researchers involved. After the communities agree on the study protocol, it has to be submitted to the process of ethical and privacy approval (if needed for the specific type of study and local legislation) and the certification of approval has to be included in the study protocol. National/Regional legislations usually require clinical trials be approved by the Data Protection Authority and/or Ethics Committee (in the U.S. by Institutional Review Boards). Other kinds of studies (e.g., observational studies) may not need any further specific and explicit approval; for example, a National/Regional legislation may allow retrospective observational studies as far as data are treated as anonymous. In any case, the communities have to assure and show in the study protocol that legal issues are respected, otherwise the TTP shall not accept to be involved in the study and the study shall be stopped. After all these steps are performed, the TTP will store the study protocol and use information within it to arrange and manage the study: e.g., it has to implement the query service to allow the data/document access, to implement the specific de-identification service for the study and arrange the patient identification management service.

Also for studies different than clinical research studies (e.g., a quality program performed by a Public Health System which need anonymous data to evaluate health services), a study protocol is needed. In this case it would not contain information specific for research studies (e.g., description of arms, IRB approval, principal investigator), however, also in this case, it has to specify information about the type of data needed, population of interest, period of analysis, de-identification technique to be used, if and how data need to be aggregated, people involved (staff authorized to perform query for data) and still the demonstration (e.g., indicating the reference legislation) about the legality if the study.

In the following section the IHE CRPC Profile is presented, which allows defining in a standard way a clinical research study protocol. It could be used also to manage the protocol of other

kinds of studies or projects (e.g., for quality purposes), even if it was not developed for this purpose.

3.1.1.1 Study Protocol Definition and CRPC Profile

The study protocol definition, in particular in the case of clinical research studies, should be compliant with the Protocol Definition Content Model, described by the Clinical Research Process Content (CRPC) QRPH Profile, which defines a clinical research study in a machine-readable format and is based on the HL7®[6] V3 Study Design Topic (RCRIM) and the CDISC Study Design Model.

According to this profile, the protocol shall contain a "Study Description" section, however it does not specify which characteristics of the study are required or not.

The study protocol should include at least the following information:

- Study ID
- Board approval information
- Study design (e.g., interventional randomized controlled trial, observational cohort study)
- Study start date
- Study closed date
- Eligibility criteria
- Data management plan (including data of interest and type of de-identification method needed)
- Outcomes
- Key staff identifiers (including e.g., researchers authorized to retrieve data)
- Protocol definition version.

In the solution we propose, the CRPC Content Creator is played by the TTP and both the Responding and Requesting Communities play the role of CRPC Content Consumer (Figure 3.1.1.1-1). Theoretically also the Communities might play the role of the CRPC Content Creator but our choice is due to the mandatory grouped actors defined by the CRPC Profile: the CRPC Content Creator has to be grouped with the RPE Process State Manager and the CRPC Content Consumer has to be grouped with the RPE Process Activity Executor. In 3.1.4 the process of patient enrollment using the RPE Profile is described: this process requires the RPE Process State Manager be played by the TTP, which is in charge to manage the patient identifiers. Therefore in our solution is the TTP that plays the role of CRPC Content Creator. It implies that even if the two communities agree on the protocol, finally it's the TTP that creates the protocol according to the CRPC Profile.

[6] HL7 is the registered trademark of Health Level Seven International.

According to the CRPC Protocol Definition Content Model, each study should have from one to multiple Study IDs. However, one Study ID should be used in order the TTP can identify the study and manage its characteristics using the Study ID as key. Each community can perform multiple studies, but, according to our model (as described in Section 2.1), in all the different studies only one patient identifier shall be used for the same patient (if different studies are performed by the same organization and need different patient identifiers, as in different clinical trials, conceptually the organization shall be considered composed of many communities, one for each patient identifier): in this case the patient enrollment (described later inSection 3.1.4) is performed just when the first study is performed.

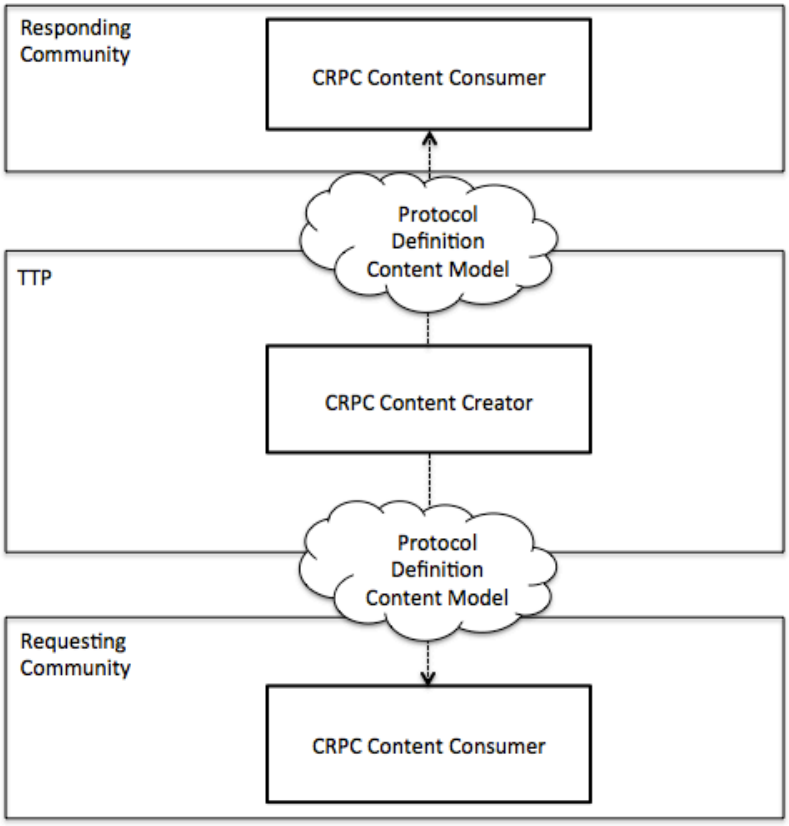


Figure 3.1.1.1-1: CRPC Protocol Definition Content Module

The definition of “Key staff identifiers” is important in order to know people allowed to perform queries and this information will be used as indicated in the Access Control white paper and described in Appendix B.

Any other required functionality of the CRPC Content Creator and Content Consumer actors about the patient’s enrollment (the “Initiate Process Content Module”) is described in Section 3.1.4.

3.1.2 Cross-community Data and Document access and Provision

In order to allow document access from a community to another one without any constraint about how a Community is organized (as a Document Sharing Affinity Domain or not), the XCA is the suitable profile to be implemented.

Since the architecture proposed in this white paper is based on the central TTP acting between the two communities and performing functionalities related in particular to document de-identification, cross-community patient management and semantic services, the XCA communication is not directly between the Requesting Community and the Responding Community, but between the TTP and the two communities.

Two different technical solutions are presented in this section, related to two main use-cases:

a) request for documents related to specific patients;

b) request for data related to specific patients.

In Section 4, other technical solutions are presented in order to manage other types of request from the Requesting Community.

3.1.2.1 Cross-Community Document Access and XCA Profile

The description of an IHE standard architecture for document access involving a central TTP, which plays the role of a bridge in the communication between the Requesting Community and the Responding Community, is represented in Figure 3.1.2.1.2-1.

An example of request for documents for specific patients can be during a clinical trial performed by a Research Organization (Requesting Community). During clinical trials, lots of data is collected through questionnaires administered to patients enrolled in the study. However, other data can be retrieve from clinical documents already collected in the Clinical Community (Responding Community), for example Discharge Summaries or Laboratory Reports. The solution here presented is intended to allow the Research Community to retrieve these kinds of documents for patients enrolled in the study without the patient's identity being disclosed to the Research Community: the patient's direct and indirect identifiers are replaced with the trial ID used in the clinical trial.

3.1.2.1.1 [ITI-38] Cross Gateway Query

The Initiating Gateway in the Requesting Community starts a [ITI-38] (Cross Gateway Query) transaction to the Responding Gateway in the TTP in order to define the query parameters (DocumentEntry metadata) and get the registry entries (documents uniqueId) about documents matching the query criteria. The registry entries will be then sent to the Responding Gateway with a [ITI-39] (Cross Gateway retrieve) transaction in order to retrieve the documents of interest.

In the [ITI-38] transaction sent by the Requesting Community, the patientId DocumentEntry metadata attribute is valued with the patient's identifier used in the Requesting Community and for whom documents need to be retrieved.

Other DocumentEntry metadata allow defining query parameters about the type of documents of interest and they are: the classCode, typeCode, formatCode and eventCodeList DocumentEntry metadata. The technical solution presented in Figure 3.1.2.1.2-1 does not show how the Requesting Community can get the information about the value of these metadata for the documents in the Responding Community storing the data of interest: it will be described in Section 3.1.5.1 where the TTP semantic service management is presented.

The TTP semantic service provides also the homeCommunityId of the Responding Community to which the other DocumentEntry metadata (classCode, typeCode, etc..) relate to. The homeCommunityID DocumentEntry metadata shall be valued with the homeCommunityID of the Responding Community of interest. According to the XCA Profile, the homeCommunityID is optional if the request is done for specific patients (patientID DocumentEntry metadata), therefore it can be omitted if the request is for specific patients and for any specific Responding Community.

Other parameters of interest for the query (e.g., the time of document creation) can be defined through other DocumentEntry metadata (the complete list of DocumentEntry metadata is defined in ITI TF vol.3).

The XCA Initiating Gateway in the TTP then forwards the request with another [ITI-38] transaction to the Responding Gateway played by the Responding Community(ies) of interest. However, before forwarding the request, the TTP has to replace in the patientId DocumentEntry metadata the patient's identifier in the Requesting Community with the patient's identifier used in the Responding Community (if the two communities use different identifiers). The management of patients' identity by the TTP and mapping of patient's identifiers are described in Section3.1.4.

If the Responding Community is organized as an XDS Affinity Domain the Responding Gateway sends a [ITI-18] transaction to the XDS Document Registry, otherwise other mechanisms have to be implemented in the Responding Community in order to retrieve the information of interest.

The [ITI-38] Response with information about the registry entries is then sent by the XCA Responding Gateway in the Responding Community to the XCA Initiating Gateway in the TTP and then forwarded to the XCA Initiating Gateway in the Requesting Community.

3.1.2.1.2 [ITI-39] Cross Gateway Retrieve

Once the XCA Initiating Gateway of the Requesting Community has received the documents uniqueId with the [ITI-38] transaction, it starts a [ITI-39] Cross Gateway Retrieve transaction to the Responding Gateway in the TTP. The request is then forwarded with another [ITI-39] transaction from the XCA Initiating Gateway in the TTP to the XCA Responding Gateway in the Responding Community(ies) of interest. If the Responding Community is organized as an XDS Affinity Domain, its Responding Gateway sends a [ITI-43] transaction to the XDS Document Repository to retrieve documents of interest; otherwise, other mechanisms have to be implemented in the Responding Community in order to retrieve them.

The XCA Responding Gateway in the Responding Community sends the [ITI-39] Response with the documents of interest to the XCA Initiating Gateway in the TTP. If for privacy reasons the patient's identity in the Responding Community cannot be disclosed to the Requesting Community, documents need to be de-identified by the TTP before being provided to the Requesting Community. The description the documents de-identification process using IHE standard solution is described in Section3.1.3.1. The documents are finally provided to the Requesting Community in the [ITI-39] Response.

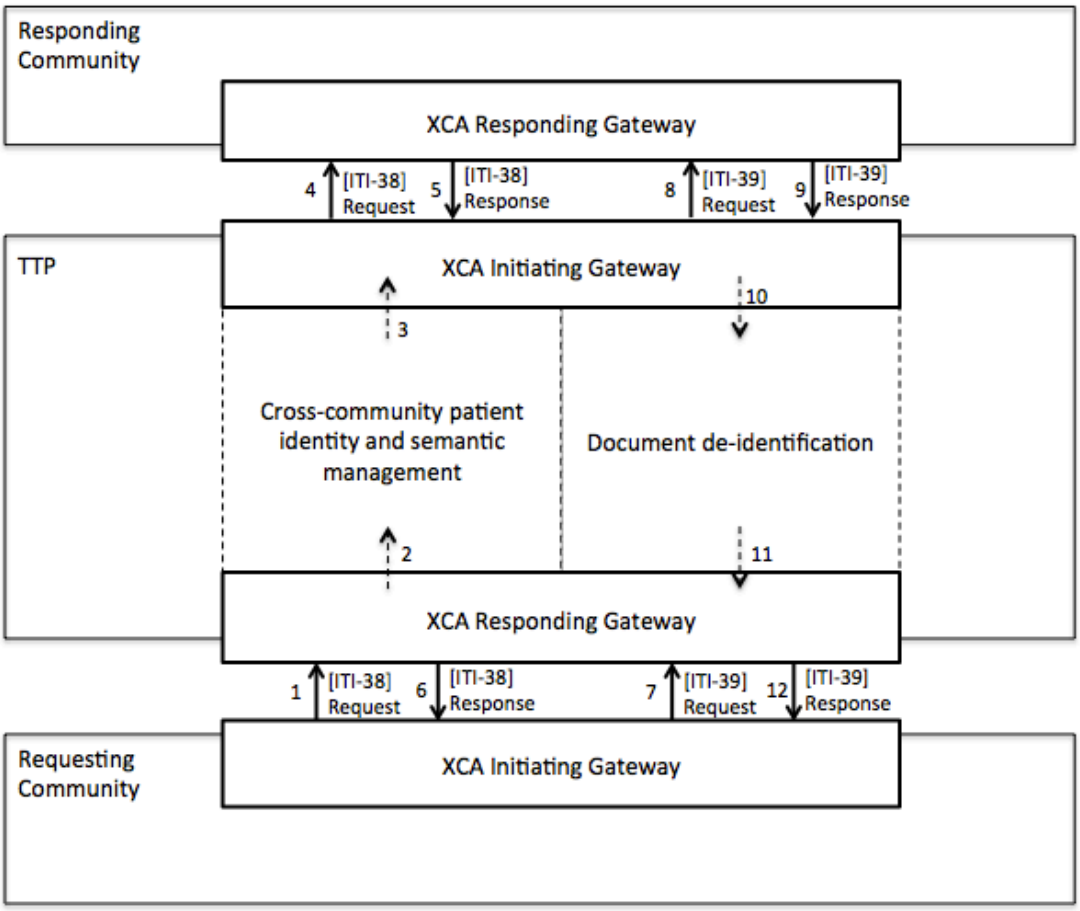


Figure 3.1.2.1.2-1: Cross-Community Document Access and XCA Profile

3.1.2.2 Cross-Community Document Access and XCDR Profile

In some scenarios it would be more suitable to implement a "push" solution instead of a "pull" solution: in these cases documents are pushed from the Responding Community to the TTP, which de-identifies documents before providing them to the Requesting Community. The "Responding Community" and "Requesting Community" terms are not very appropriate for this

context because it is a “push” solution without any requests for data, however the same terminolgy is used for coherence with the other scenarios: the “Responding Community” is the community providing data, the “Requesting Community” that receiving data. A use case when this solution may be implemented is a Regional health authority (Requesting Community) wanting to be informed every day for governance purposes and costs control about drug prescriptions created in a health information exchange (HIE) system (Responding Community): in this case everytime a drug prescription is created in the HIE, it is sent to the TTP, where it is de-identified and finally provided to the Regional health authority.

In this scenario the suitable IHE profile to be applied is the Cross-Community Document Reliable Interchange (XCDR) Profile, which specifies the [ITI-80] Cross-Gateway Document Provide transaction to push document from the Initiating Gateway of a source Community to the Responding Gateway of a target Community (see Figure 3.1.2.2-1). This transaction provides documents and contains metadata that allows the Responding Gateway to process the documents. In the study protocol, the type of documents to be sent, the patients (or eligibility criteria defining patients of interest) for which documents need to be sent, period of time of interest about document creation, and modality of document transmission (e.g., synchronous or asynchronous web service exchange) have to be defined. When the TTP performs the document de-identification (anonymization or pseudonymization with replacement of the patient’s identifier of the first community with the patient’s identifier of the second community), also metadata need to be de-identified before starting the second [ITI-80] transaction to the final community.

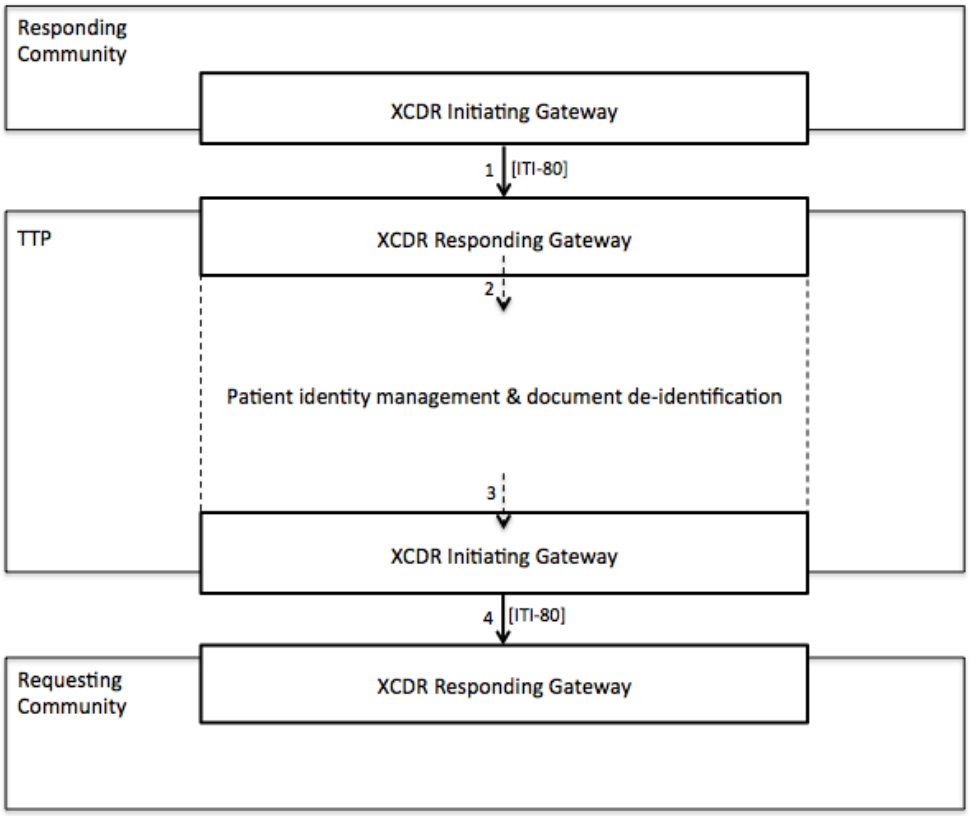


Figure 3.1.2.2-1: Cross-Community Document Access and XCDR Profile

3.1.2.3 Cross-Community Data Access and XCA and QED Profiles

Figure 3.1.2.3.1-1 shows an IHE standard architecture for data access involving a central TTP playing the role of a bridge in the communication between the Requesting Community and the Responding Community.

An example of request for data for specific patients is a clinical trial performed by a Research Organization (Requesting Community). During clinical trials lots of data are collected through questionnaires administered to patients enrolled in the study. However other data can be retrieved from clinical documents already collected in the Clinical Community (Responding Community), for example Discharge Summaries or Laboratory Reports. The solution here presented is intended to allow the Research Community to retrieve data of interest (e.g., patient's diagnosis, vital signs, laboratory results) for patients enrolled in the study without disclosing the patient's identity to the Research Community.

The standard solution for data access is similar to that for document access, however in this case the communication between the Requesting Community and the TTP does not comply the XCA Profile but the Query for Existing Data (QED) Profile since the request is not for document access but for data access.

3.1.2.3.1 [PCC-1] Query for Existing Data

A [PCC-1] Request is sent from a QED Clinical Data Consumer in the Requesting Community to the QED Clinical Data Source in the TTP. This transaction allows to asks for clinical data for specific patients, in particular to the following data categories:

- Vital signs (simple measurements or reported values that can be determined using simple measuring devices (e.g., Height, Weight), or which can be reported by the patient (date of last menstrual period))
- Problems and allergies (diagnoses, clinical findings, allergies, or other risk factor)
- Diagnostic results (observations made or performed using laboratory testing equipment, imaging procedures, vision examinations, etcetera)
- Medications (medications that a patient is or has been taking for treatment of one or more conditions)
- Immunizations (immunizations that have been given, or which are planned to be given to a patient)
- Professional services (procedures and/or encounters which the patient has participated in, or is expected to participate in)

In order to be able to provide the different types of data, both the QED Clinical Data Consumer and the QED Clinical Data Source have to implement the option corresponding to the kind of data category of interest.

The QED Clinical Data Consumer is allowed to ask data for specific patients indicating the patient's identifier used in the Requesting Community within the patientID parameter in the [PCC-1] Request. A request can be performed also for multiple patients (the technical solution is described in Section 4.1.1) and/or be provided as aggregate data (the technical solution is described in Section 4.2). The QED Clinical Data Source in the TTP has to be grouped with actors managing the patient's identity: in this way the patient's identifier of the Requesting Community is replaced with the patient's identifier used in the Responding Community (if the two communities use different identifiers). Other filter parameters can be defined in the [PCC-1] Request, as the effective time for the clinical statement, which is provided within the clinicalStatementTimePeriod parameter.

The QED Clinical Data Consumer does not have to specify the Responding Community of interest in the [PCC-1] Request, because the TTP knows the information about all the Responding Communities involved in the study to which forward the request. Moreover, the QED Clinical Data Source in the TTP should be grouped with actors providing semantic information (as described in Section3.1.5.1) allowing to identify the Communities storing the data of interest: only to these Responding Communities the request for documents retrieval is forwarded. The semantic service provides also the information about the Document Entry Metadata (classCode, typeCode, formatCode and eventCodeList DocumentEntry metadata) describing the types of documents of interest.

The TTP performs then a [ITI-38] Cross Gateway Query transaction Request in order to get the registry entries (documents uniqueId) about documents matching the criteria indicated by the Requesting Community in the [PCC-1] Request. The [ITI-38] Cross Gateway Query transaction Request is sent by the XCA Initiating Gateway in the TTP to the XCA Responding Gateway of the Responding Community. The filter parameters of the request are provided within the Document Entry Metadata: they allow to describe the types of documents of interest (classCode, typeCode, formatCode and eventCodeList), the patient (patientId), the community of interest (homeCommunityId), the time of document creation (serviceStartTime and serviceStopTime), etc. The Response of the [ITI-38] transaction provides the documents uniqueId useful to initiate a [ITI-39] Cross Gateway Retrieve transaction. In the [ITI-39] Response the Responding Community provides to the TTP the clinical documents of interest. The TTP then performs the extraction of clinical data of interest from documents and it returns them to the QED Clinical Data Consumer in the [PCC-1] Response. In the [PCC-1] Response the patient's identifier is that used by the Requesting Community.

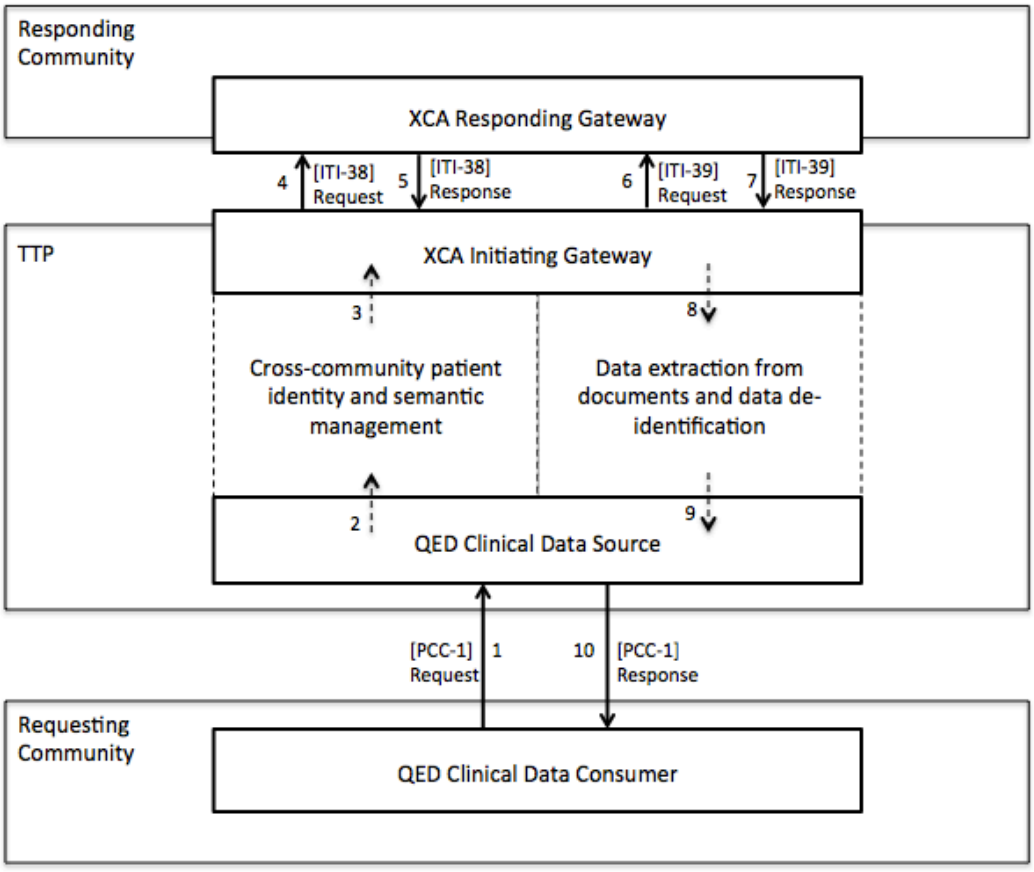


Figure 3.1.2.3.1-1: Cross-Community Data Access and QED Profile

3.1.3 Cross-Community De-identification Service

The core functionality of the TTP is the de-identification activity, as described in the Section 2.4 Privacy and Security Considerations. In the study protocol, also the type of de-identification technique has to be defined, (anonymization, pseudonymization, aggregate data, or none of them).

In particular, for each type of document of interest, the following information should be defined:

- Which data (direct identifying data) need to be replaced with a pseudonym (if pseudonymization is needed to allow longitudinal consistency and relinking) or an anonymous identifier (not relinking is needed);
- Which data need to be processed and the technique used (for example fuzzing or generalization);
- Which data need to be preserved.

All other data (direct and indirect identifying data and not necessary data) need to be removed (redaction algorithm).

The classification indicated above should be performed according to the principles of relevance, adequacy, efficiency and not excess. It means that non-necessary data should be removed (principle of non-excess), that the technique adopted should be efficient to protect the patient's identity but in the meantime adequate and relevant for the purpose of the study.

As indicated in the glossary, it is important to specify that in this white paper with the term "pseudonym" we refer to the patient identifier used by the Requesting Community that is in general different to that used by the Responding Community. Therefore, the term "pseudonym" is used in all the following situations:

a) The patient identifier (pseudonym) used in the Requesting Community is created ad hoc at the beginning of the study as corresponding to the patient identifier used in the Responding Community. In this case the pseudonym is not only created for de-identification purposes but also because a patient identifier in the Requesting Community does not yet exist (e.g., this is the case of the beginning of a clinical trial when a patient belonging to a Clinical Community is enrolled in a clinical trial performed by a Research Community and a Study ID for that patient in the Research Community is created);

b) The patient identifier (pseudonym) already exists in the Requesting Community at the beginning of the study: when the study starts, a linkage between the patient identifier used in the Responding Community and the pseudonym is created (e.g., this is the case of a Research Community that has already enrolled the patient in a study and now wants to have access to other patient's data available in a Clinical Community);

c) The patient identifier (pseudonym) used in the Requesting Community does not exist at the beginning of the study but it is created ad-hoc the first time one of his/her documents/data need to be de-identified before being provided to the Requesting Community: the pseudonym replaces the patient identifier used in the Responding

Community and becomes the patient identifier used in the Requesting Community. For example this is the case of a epidemiology institution which is interested to have access to patients' summary from a clinical community in order to find people with a specific disease and then follow them during time in a longitudinal study to evaluate the complications of the disease: every time a patient summary is created in the clinical community, it is pushed to the TTP (as indicated in Section 3.1.2.2) which creates a pseudonym and pushes again the pseudonymized document to the epidemiology institution.

With the term "pseudonymization" we refer in general to the replacement of the patient identifier used in the Responding Community with the corresponding "pseudonym" used in the Requesting Community.

The de-identification service is performed using IHE actors from the Redaction Services (RSP) Profile and Patient Identifier Cross-Referencing (PIX) Profile.

3.1.3.1 Document De-Identification and RSP - PIX Profiles

The RSP Profile provides a method to redact data from a document according to an extraction specification provided by an external system. Therefore, this profile allows to process documents in order to de-identify documents according to specific rules defined as extraction specifications (as an XSLT). The de-identification process happens after documents are retrieved from the Responding Community and before they are delivered to the Requesting Community. Extraction specifications can indicate an actual reduction of data as well as a data elaboration (e.g., date generalization: the YYYYMMDD date format is replaced by the YYYY date format). The result of the application of this profile is a redacted document. In use cases where the Requesting Community is interested on de-identified data instead of de-identified documents, a process of data extraction from documents will follow (as described in Section 3.1.2.3). Details about the process of document request and retrieval are not shown in the picture (for details see Section3.1.2).

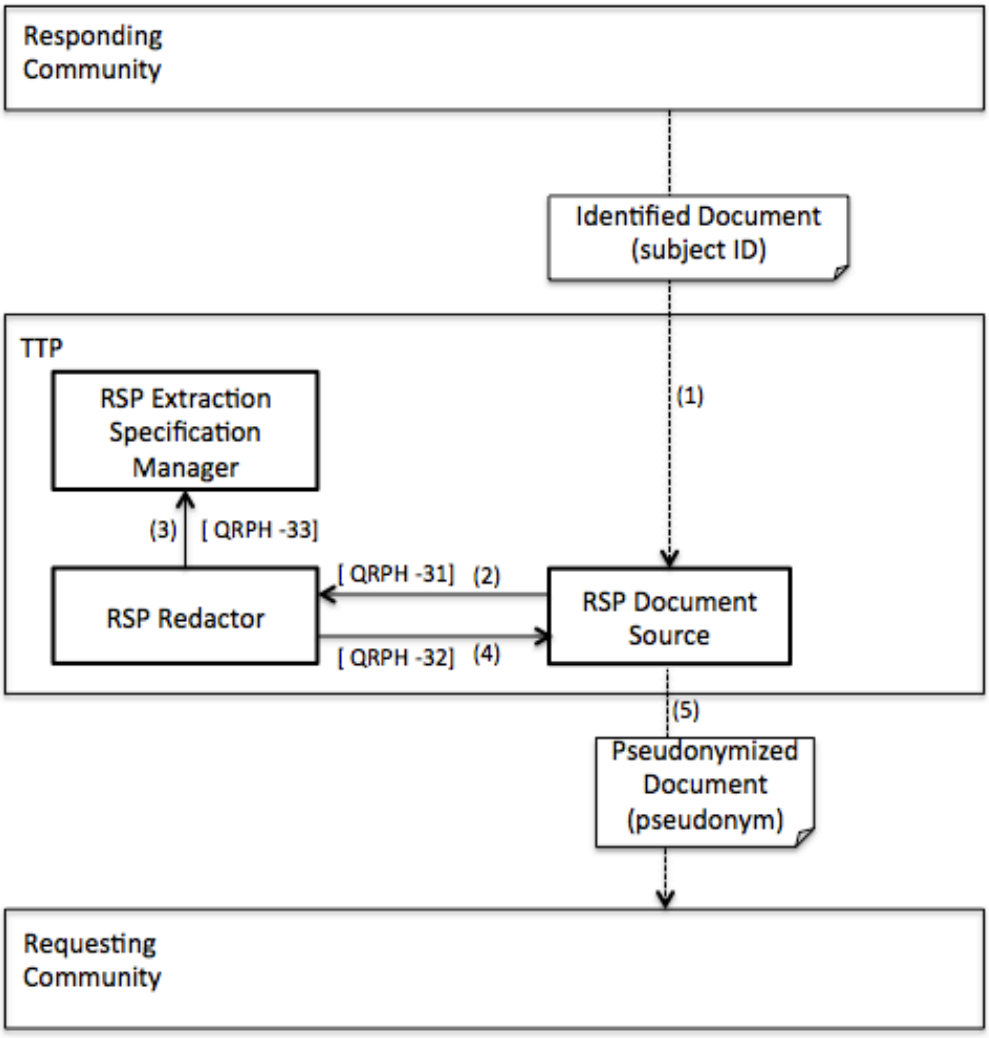


Figure 3.1.3.1-1: De-identification and RSP Profile

The RSP Profile alone does not allow performing the pseudonymization technique; however, in conjunction with PIX Profile, also this functionality can be provided (Figure 3.1.3.1-2). The pseudonymization process can be seen as the combination of creation/retrieval of the pseudonym, redaction of direct identifying data, in particular blanking of the main patient identifier used by the Responding Community and its replacement with the pseudonym used by the Requesting Community (the process of pseudonym creation is described in Section 3.1.4). The PIX actors involved in the pseudonymization are the Patient Identifier Cross-reference Consumer and the Patient Identifier Cross-reference Manager. The Patient Identifier Cross-reference Consumer asks with the [ITI-9] transaction to the Patient Identifier Cross-reference Manager which is the identifier (pseudonym) used by the Requesting Community corresponding to the identifier used by the Responding Community. After this information is obtained, the

identifier used by the Responding Community is blanked and other direct identifying data are redacted using the RSP Profile (the PIX Patient Identifier Cross-reference Consumer should be grouped with the RSP Document Source). Finally, the patient's identifier element in the document (previously showing the identifier used by the Responding Community) is filled-in with the pseudonym.

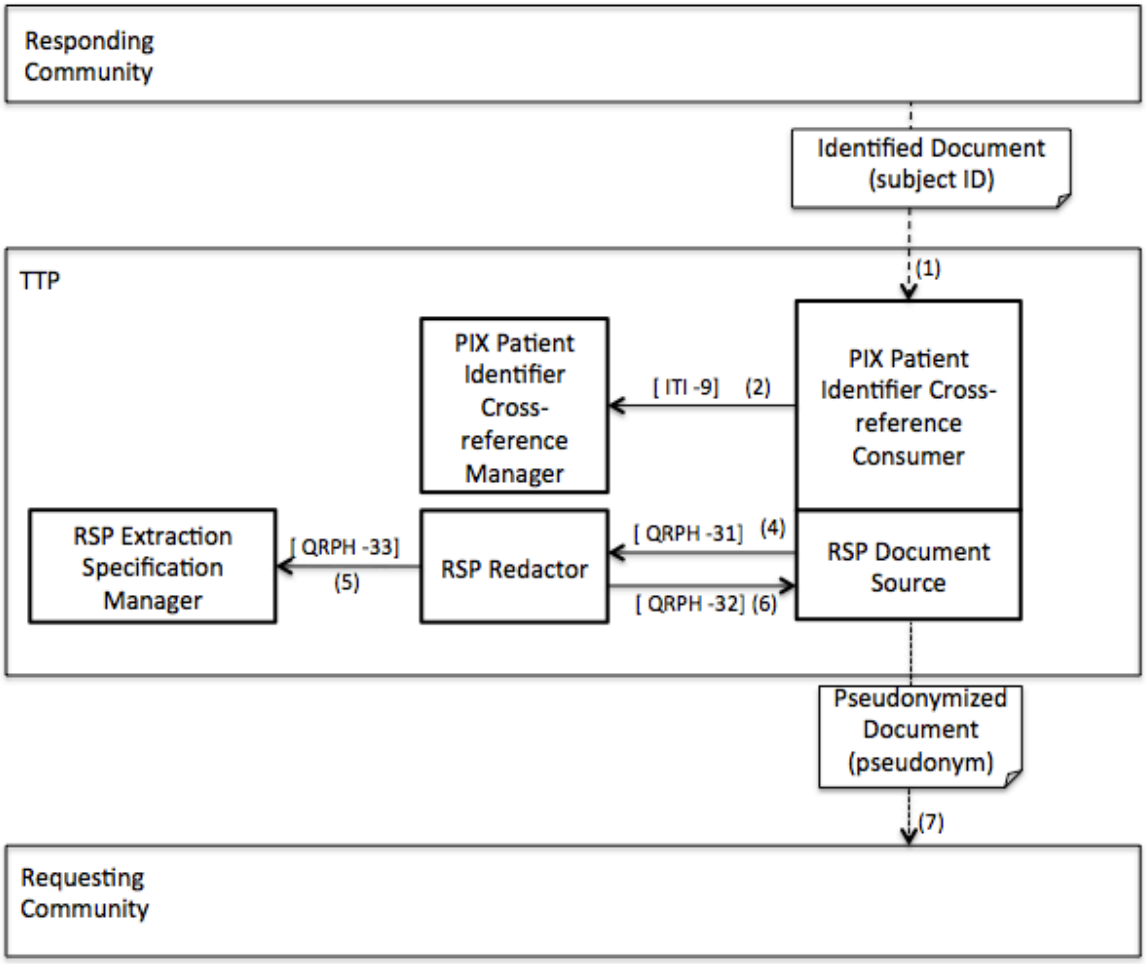


Figure 3.1.3.1-2: Pseudonymization and RSP and PIX Profiles

3.1.3.2 Document De-Identification if TTP has forbidden access to clinical data

If a specific jurisdiction does not allow the TTP to have access to clinical data, but only to the PII, in order to perform the document pseudonymization, a technical solution different to those presented so far has to be adopted. Here an overview of the design of this technical solution is presented, however it would involve some actors and transactions not already defined by IHE.

Figure 3.1.3.2-1 shows this solution and highlights (with dotted lines) the no-IHE-compliant actors and transactions. The idea behind this solution is that the original identified document has to be split into two parts, one containing PII and the other containing the patient's clinical data. Moreover this solution takes into account only a "push" mechanism initiated by the Responding Community, which wants to provide to the Requesting Community the patient's documents as pseudonymized documents (in case of anonymized documents the solution is simpler and involves only the RSP actors and transactions). A "pull" mechanism (with an actual Request by the Requesting Community to have access to a specific document) is even a more complicated case, because neither XDS nor XCA, nor any other IHE profiles can be applied involving such a mechanism for documents splitting and reconstruction.

As illustrated in Figure 3.1.3.2-1, the process starts with two parallel transaction flows:

- The first one (a) aimed to redact a document in order to obtain a document only with clinical data of interest for the Requesting Community;
- The second one (b) aimed to save the patient's identifier used in the Responding Community (Subject ID) and to identify the correspondent patient's identifier used in the Requesting Community.

The a) flow starts with a Provide and Register Document Set-b [ITI-41] transaction sent by the XDR Document Source to the XDR Document Recipient (both the two actors played by the Responding Community). With this transaction the identified document, containing both PII and clinical data, is sent to a RSP Document Source grouped with the XDR Document Recipient. The document identifier is the IDdocOrig and it is provided in the [ITI-41] transaction within the uniqueID Document Entry metadata: this information is the key for all following steps and will be used at the end of the process to merge all the split information. Therefore, the IDdocOrig shall be shared by the XDR Document Recipient to all the grouped actors in order to keep track of the original document identifier. The document is redacted by RSP actors through the [QRPH-31], [QRPH-32] and [QRPH-33] transactions. All the RSP actors are played by the Responding Community and this is one of the main difference in comparison to the general solution presented in Section3.1.3.1, where the redaction service is performed by the TTP. The redacted document containing only the clinical data is finally sent by the XCDR Initiating Gateway in the Responding Community to the XCDR Responding Gateway in the Requesting Community through the [ITI-80] transaction. Since the redacted document in this case does not relate to any specific patient, the patientID Document Entry metadata shall be valued with a wildcard agreed by the two communities and meaning that the document is not related to any patient. Also the IDdocOrig information shall be conveyed through this transaction: a possible solution is to identify the redacted document with the same identifier of the original document (IDdocOrig), another one is to provide this information within the "referenceIdList" metadata of the [ITI-80] transaction.

The b) flow starts at the same time of flow a) and in the first step the Subject ID stored within the identified document is sent to the PIX Patient Identifier Cross-Reference Consumer together with the IDdocOrig. This transaction (1b) is not IHE compliant, because no IHE transactions allow providing both the subject ID and the document ID. The PIX Patient Identifier Cross-Reference

Manager is then interrogated with the [ITI-9] transaction in order to get the pseudonym used by the Requesting Community and correspondent to the subject ID used by the Responding Community. Another transaction not IHE compliant is sent then by the PIX Patient Identifier Cross-Reference Consumer to provide the pseudonym together with the IDdocOrig to a Document Creator played by the Responding Community (3b).

Finally, the Document Creator, which is grouped with the XCDR Responding Gateway, merges the pseudonym with the correspondent redacted document (they have the same IDdocOrig) and creates the final pseudonymized document.

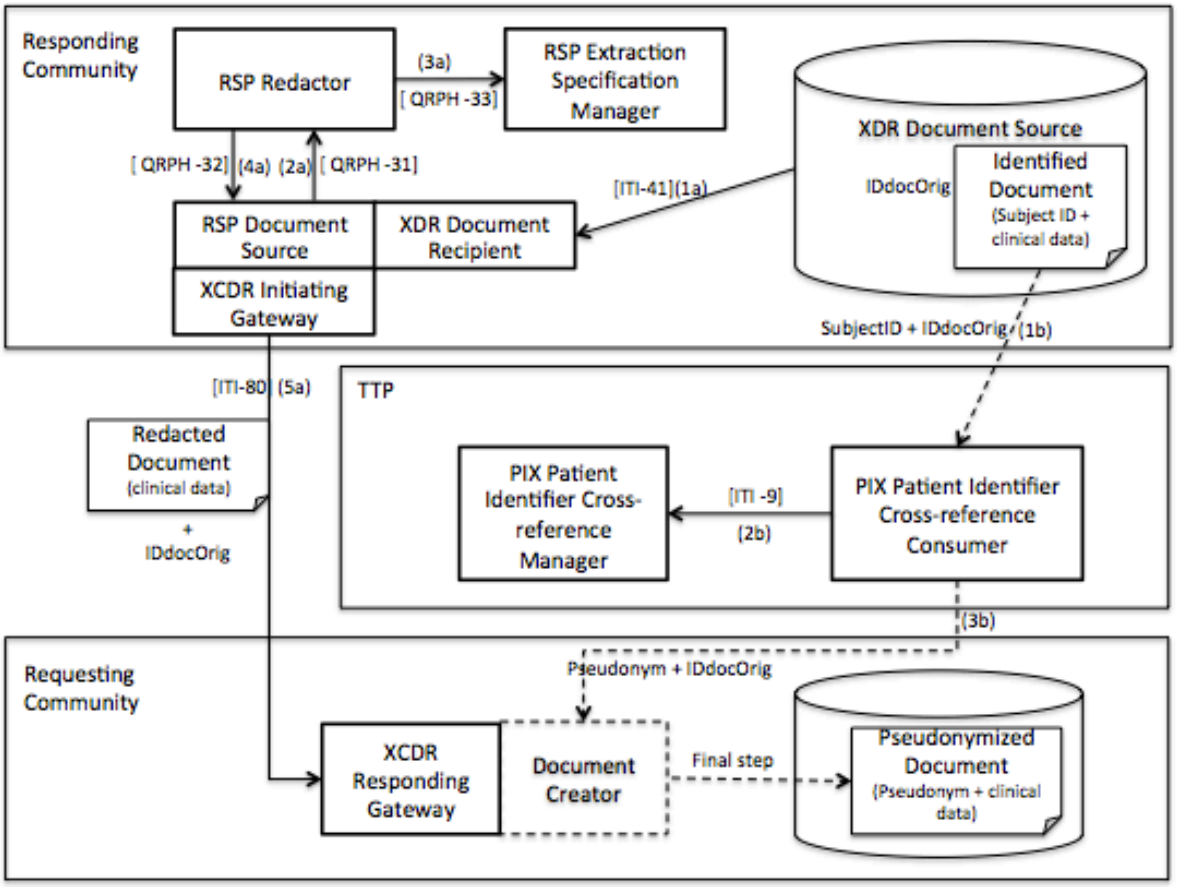


Figure 3.1.3.2-1: Overview of technical solution if TTP has forbidden access to clinical data

3.1.4 Cross-Community Patient Identity Management

As discussed so far, another important functionality performed by the TTP is the patient identity management, which involves, first of all, the management of correspondences between the

patient identifiers used by the Requesting Community and those used by the Responding Community. This is a required functionality in use-cases needing the "pseudonymization", so when two different identifiers have to be used in the two different communities, as described above in Section 3.1.3. Sometimes even for the anonymization technique, a new anonymous identifier might need to be created: however, in this case, the management and storing of correspondences between identifiers is not required and specific actors to perform these activities are not needed.

3.1.4.1 Patient Identity Management and PIX and RPE Profiles

The PIX Patient Identifier Cross-reference Manager performs the management of correspondences between the patient identifiers used in the different communities. Section 3.1.3.1 describes how this actor can be queried to get information about the correspondences between patient identifiers. This section focuses on the definition of patient identifiers and managing of correspondences between them in order to populate the PIX Patient Identifier Cross-reference Manager.

Three different use-cases can be identified according to the three "pseudonym" usages presented in the introduction of Section 3.1.3.

The three technical solutions related to each specific use-case are proposed here below: the technical solution is indicated with the same letter of the correspondent use-case in Section 3.1.3. The assumption, which is in common to all the three solutions, is that only the TTP can know the association between patient identifiers used in the different communities and therefore it plays the role of the PIX Patient Identifier Cross-reference Manager.

a) The patient identifier (pseudonym) to be used in the Requesting Community is created ad hoc by the TTP at the beginning of the study as corresponding to the subject ID used in the Responding Community. In this case, the pseudonym is not only created for de-identification purposes but also because a patient identifier in the Requesting Community does not yet exist. The pseudonym has to be created in order to be unique in the Requesting Community, where each patient is identified with his/her unique pseudonym. Moreover, in general the pseudonym has to be different from the subject ID because the patient's identity in the Responding Community in general cannot be disclosed to the Requesting Community. If the same patient's identifier is allowed in the two communities, it shall be explicitly stated and demonstrated in the study protocol, as described in Section 3.1.1. The protocol should establish also the rules about the creation of the pseudonym, for example about its length, its format (numeric/ alphanumeric/ text), specific characters to be used.

   A typical use-case is the enrolment of a patient belonging to a Clinical Community into a clinical trial performed by a Research Community and a pseudonym for the Research Community is created. In the architecture here proposed the patient's enrolment in a clinical trial is managed by the TTP, where the patient is registered and where the pseudonym is generated.

The TTP plays the role of the Process State Manager, under the RPE (Retrieve Process for Execution) Profile. Both the Requesting and Responding Community play the role of the Process Activity Executor actors, under the RPE Profile.

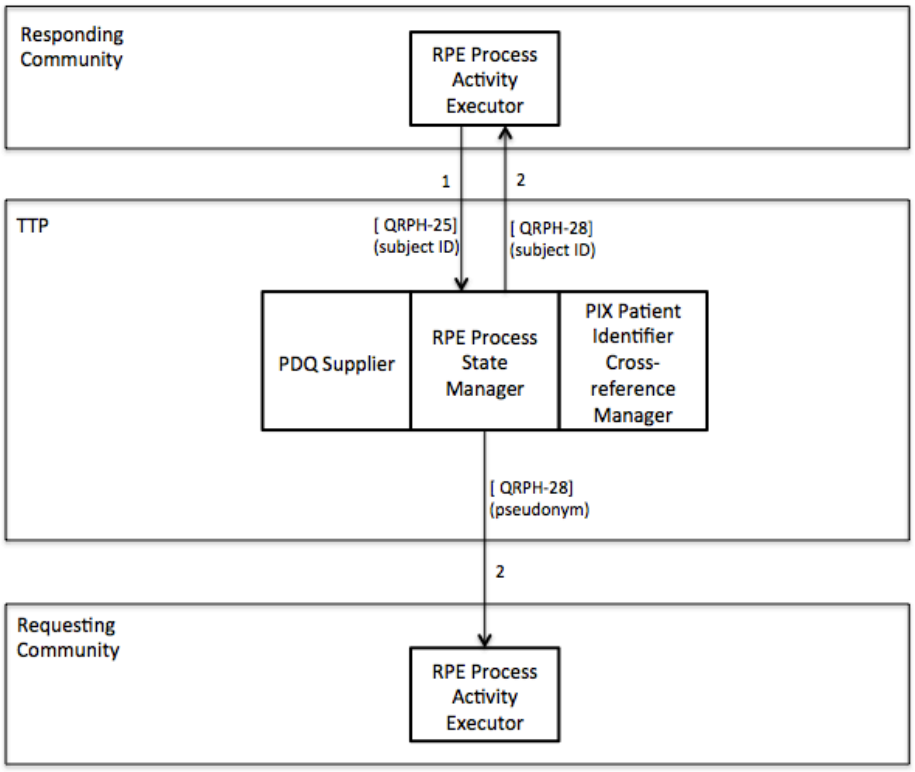


Figure 3.1.4.1-1: Patient enrolment and RPE and PIX profiles (case a)

In Figure 3.1.4.1-1, the process of the patient enrollment using IHE profiles is described. In our model the community where the patient is recruited is the "Responding Community" because this is the community where documents belonging to the patient already exist (in the other community documents cannot already exist because at the beginning of the study the community does not have any patient). The patient identifier in the Responding Community is the "subject ID" in the figure. With the Initiate Process [QRPH-25] transaction, the subject ID is sent by the RPE Process Activity Executor in the Responding Community to the RPE Process State Manager, which enrolls the patient in the study and creates the pseudonym to be used as the patient identifier in the Requesting Community. The Initiate Process [QRPH-25] transaction is usually performed after the [QRPH-20] Retrieve Process Definitions transaction, which allows the RPE Process Activity Executor to know the activities to perform and so the process to initiate. However, in this case, this preliminary step may not be necessary because the RPE Process Activity Executors already know about enrollment and the other activities

to be performed, since the study protocol (defined in compliance with the CRPC Profile, Section 3.1.1.1) has already been agreed and shared by the communities and TTP. The RPE Process Activity Executor can send also the patient's demographic characteristics with the [QRPH-25] transaction (they are optional data in the [QRPH-25] Initiate Process Request): they will be stored by the PDQ Supplier grouped with the RPE Process State Manager. The notification of the enrolment is sent to the RPE Process Activity Executor in the Requesting Community (with the Send Process State Alert [QRPH-28] transaction containing the "pseudonym"). The pseudonym will then be used by the Requesting Community to ask for data/documents related to the patient identified by the pseudonym. An optional [QRPH-28] transaction can be sent also to the RPE Process Activity Executor in the Responding Community (with the Send Process State Alert [QRPH-28] transaction containing the "subject ID") in order to notify the Responding Community the enrolment of the patient was performed. The PIX Patient Identifier Cross-reference Manager is grouped with the RPE Process State Manager in order to store and manage the correspondence between the subject ID and the pseudonym for the two specific communities.

b) The patient identifier (pseudonym) already exists in the Requesting Community at the beginning of the study: when the study starts, a linkage between the patient identifier used in the Responding Community and the pseudonym is created. Two technical solutions are here presented to manage this situation.

b1) The first solution is similar to that presented for case (a). However,, in this case the request for the patient's enrolment is performed both by the RPE Process Activity Executor of the Responding Community (with transaction [QRPH-25] providing the subject ID) and the RPE Process Activity Executor of the Requesting Community (with transaction [QRPH-25] providing the pseudonym) (Figure 3.1.4.1-2). Also in this case, the Initiate Process [QRPH-25] transaction does not have to be preceded by the [QRPH-20] Retrieve Process Definitions transaction: the RPE Process Activity Executors already know about the enrollment and the other activities to be performed, since the study protocol (defined in compliance with the CRPC Profile, Section 3.1.1.1) has already been agreed and shared by the communities and TTP. The two communities identify the patients to be enrolled according to the eligibility criteria defines in the study protocol (Section 3.1.1). The two requests for correspondent patients in general are performed in two different moments, because the two communities do not know (and do not have to know) which is the patient identifier correspondent in the other community. The QRPH-25 transaction is used to inform the TTP about a patient that potentially can be enrolled in the study, but the enrollment is actually performed only if and after a linkage with another patient from the other community is established by the TTP. The TTP has to implement its own (software and/or manual) solutions and algorithms to perform the linkage, which is usually a probabilistic linkage since the two communities use generally different patient identifiers for privacy issues. Patient's demographic

characteristics, which would allow performing the linkage, are sent by the RPE Process Activity Executors in the [QRPH-25] Initiate Process Request. According to the RPE Profile, the demographic data are optional in the [QRPH-25] Request, however in this specific implementation they should be defined as required, unless the TTP does not have any other further internal mechanisms to have access to this information and/or to perform the linkage. These demographic data are used by the PIX Patient Identifier Cross-reference Manager (grouped with RPE Process State Manager) to perform the linkage between the patient identifiers of the two communities related to the same patient and they are stored by the PDQ Supplier grouped with RPE Process State Manager. A notification about the successful patient enrolment is finally sent at the same time to both the two RPE Process Activity Executor actors through [QRPH-28] transactions.

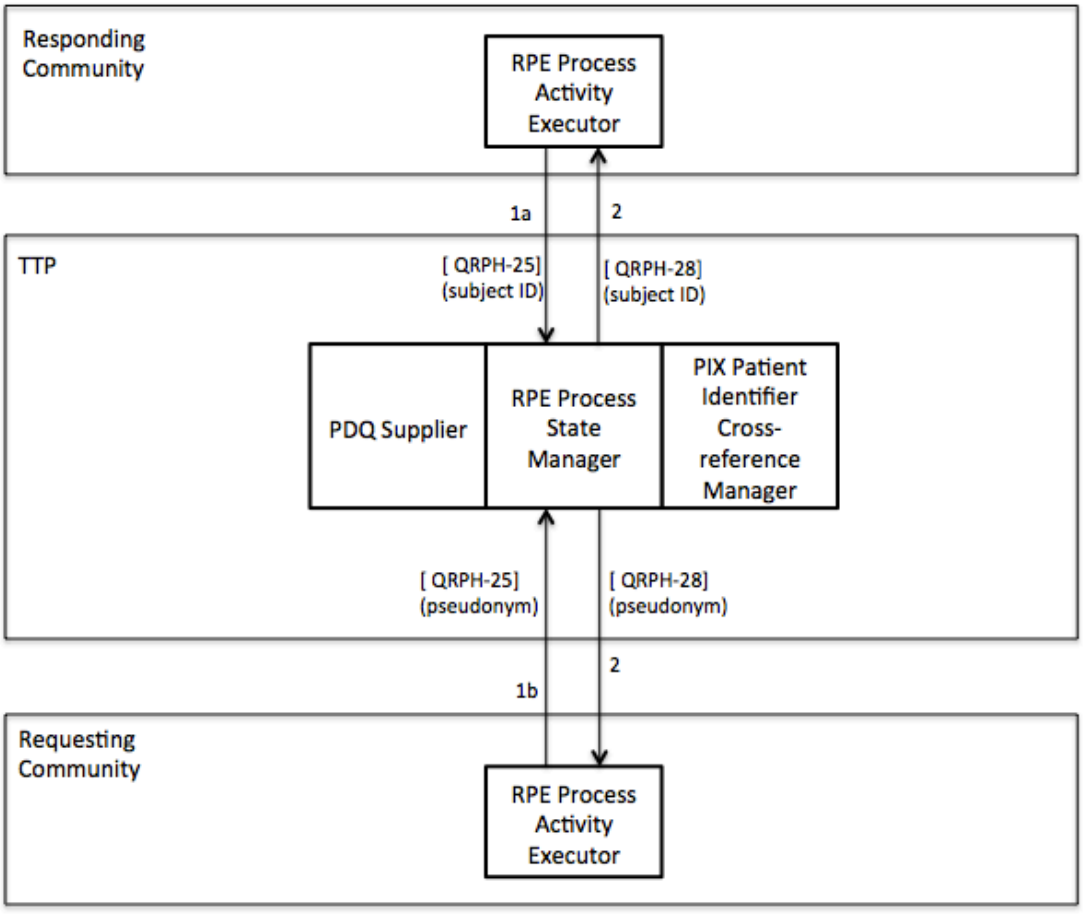


Figure 3.1.4.1-2: Patient enrollment and RPE and PIX profiles (case b1)

b2) The second technical solution involves only the PIX Profile (see Figure 3.1.4.1-3). The PIX Profile supports the cross-referencing of patient identifiers from multiple Patient Identifier Domains via the following interactions:

• The transmission of patient identity information from an identity source to the Patient Identifier Cross-reference Manager ([ITI-8] transaction).

• The ability to access the list(s) of cross-referenced patient identifiers either via a query/ response ([ITI-9] transaction) or via update notification ([ITI-10] transaction).

However, in this specific case it is applied only to manage cross-reference of patient identifiers from the different communities and not to provide to a community the patient identifier used in the other community (privacy issue). This specific requirement has some consequences that are analyzed within this section.

In this case both a PIX Identity Source in the Responding and Requesting Community send with a Patient Identity Feed [ITI-8] transaction to the PIX Patient Identifier Cross-reference Manager in the TTP the identifiers of patients eligible for the study and that may be potentially enrolled in the study if a linkage with a patient in the other community will be performed. Within the [ITI-8] transaction, the patient's demographic characteristics are provided: they allow then the PIX Patient Identifier Cross-reference Manager to perform the linkage if a matching of patients is identified. The patient name in PID-5 in the ITI-8 Request is a required field, but it is likely that, for privacy reasons, at least one of the two communities (e.g., a Research Organization) is not allowed to know the direct patient's identifiers: a workaround can be used replacing the patient's name with a pre-defined string (e.g., '99999'). The two [ITI-8] Requests for correspondent patients are necessary sent by the two communities in different moments, because they do not know (and do not have to know) which is the patient identifier correspondent in the other community. Also, in this case, the [ITI-8] transaction is used to inform the TTP about a patient that potentially can be enrolled in the study, but the enrollment is actually performed only if and after a linkage with another patient from the other community is established by the Patient Identifier Cross-reference Manager. It has to implement its own (software and/or manual) solutions and algorithms to perform the linkage, which is usually a probabilistic linkage since the two communities use generally different patient identifiers for privacy issues. The two communities usually need to know when a linkage is performed and so a patient is actually enrolled in the study. This notification can be managed in two ways. The first solution is to use the PIX Update Notification [ITI-10] transaction to the PIX Patient Identifier Cross-reference Consumers in the two communities grouped with the PIX Identity Source of each community. However, also in this case a workaround shall be adopted because this transaction requires the PIX Patient Identifier Cross-

reference Manager notifies to the communities interested in receiving notifications the list of cross-reference patient identifiers, so also the patient identifier used in the other community, which shall be avoided. Therefore, for our purposes, a notification with [ITI-10] transaction is sent only after the patient identifier used in the other community is replaced with a pre-defined string (e.g., '99999'). Another solution can be adopted to notify the communities about a successful matching, but this specific solution can be applied only if the Responding Community does not need a notification about the linkage. In this case first the Responding Community sends with multiple [ITI-8] transactions to the PIX Patient Identifier Cross-reference Manager, the lists of identifiers (together with his/her demographic data) of all its own eligilbe patients, and secondly the Requesting Community with multiple [ITI-8] transactions sends to the PIX Patient Identifier Cross-reference Manager, the lists of identifiers (together with his/her demographic data) of all its own eligilbe patients. In this way the PIX Patient Identifier Cross-reference Manager can perform the linkage as soon as a [ITI-8] Request is sent by the Requesting Community. Therefore the information about the successful/unsuccessful matching can be soon conveyed in the [ITI-8] Response (ACK=successful matching, a specific error code for "unmatched patient" to be provided in the Response shall be previously define to

the Requesting Community and no more transactions are needed.

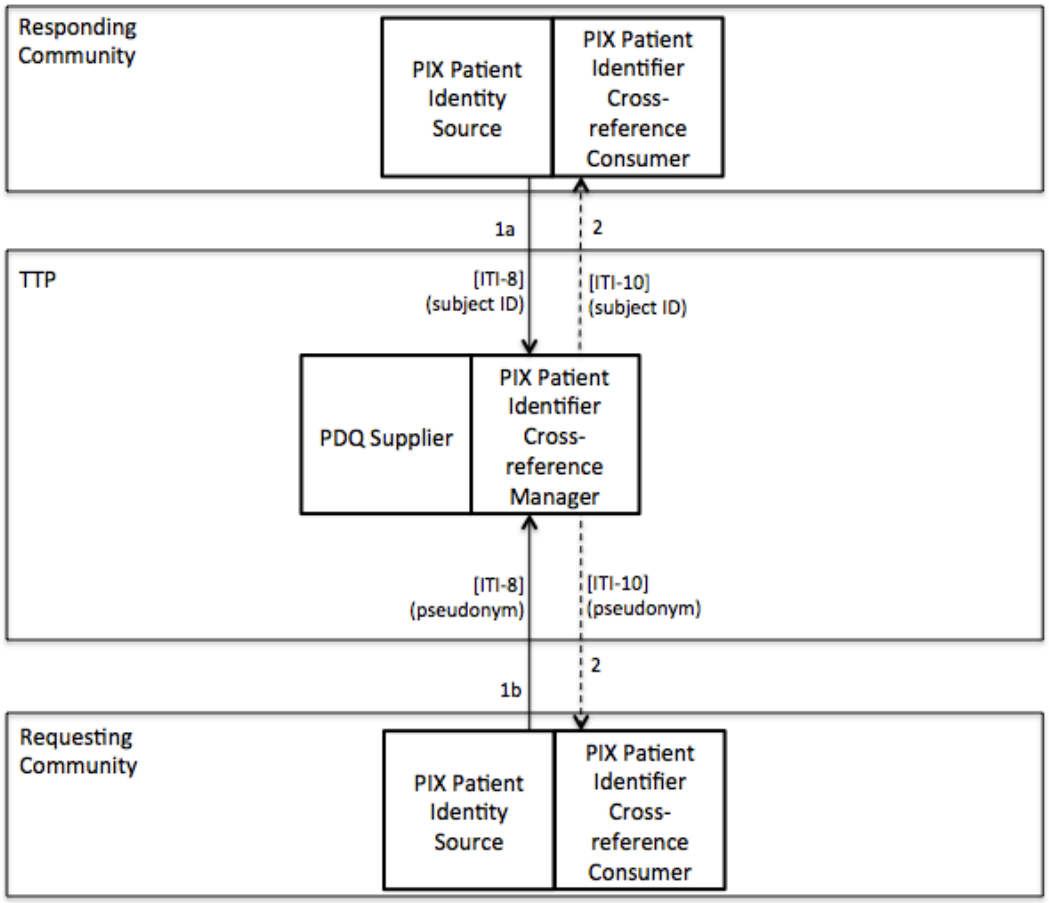


Figure 3.1.4.1-3: Patient enrollment and PIX Profile (case b2)

c) The pseudonym is created during the document de-identification process (where in this case "de-identification" means "pseudonymization"). This use case happens when the "push option" is implemented (Section 3.1.2.2), that is when documents are provided to the Requesting community not after a request for specific documents is sent, but every time they match some criteria defined in the study protocol (e.g., if the related patient is eligible or the type of document is of interest). Therefore, the "Requesting Community" does not perform an actual request with a query, but for coherence with the terminology used so far, it is still called as "Requesting" Community meaning this is the community to which documents are provided. In this case when the TTP retrieves the identified documents from the Responding Community, it extracts the subject ID (from the document itself or from metadata related to that document) and it checks with the PIX Query [ITI-9] transaction to the PIX Patient-Identifier Cross- Reference Manager (played by another system within the TTP) if the subject ID has already been recorded and associated with an identifier in the Requesting Community. If an error is returned, so

if the enrollment has not been already performed, the patient is “enrolled”, but, differently from “case a”, all the enrollment process is performed within the TTP: a RPE Process Activity Executor (played by a system in the TTP) is grouped with the PIX Patient Identifier Cross-reference Consumer and through a [QRPH-25] transaction it provides to the RPE Process State Manager (played by another system in the TTP) the subject ID related to the patient to be enrolled. In this case, it is not a real enrollment, because there is not a request for enrolment from the two communities: it is more an implicit enrollment because the patients involved in the study are determined by the TTP directly extracting this information from the document itself. It can be considered anyway such as enrollment because it happens at the first time the patient is identified from one of his/her document and since that moment the same pseudonym is used by the Requesting Community. Also in this case, the Initiate Process [QRPH-25] transaction is not preceded by the [QRPH-20] Retrieve Process Definitions transaction, because the RPE Process Activity Executor already know about enrollment and the other activities to be performed, since the study protocol (defined in compliance with the CRPC Profile, Section 3.1.1.1) has already been agreed and shared with the TTP. The RPE Process State Manager enrolls the patient, generates the pseudonym to be used in the Requesting Community, and return the enrolment notification with the pseudonym to the RPE Process Activity Executor with the [QRPH-28] transaction. In the meantime, the PIX Patient Identifier Cross-reference Manager stores the association between the two identifiers. An optional PDQ Supplier can be grouped with the PIX Patient Identifier Cross-reference Manager in order to store demographic information available in the documents or in metadata. After this “enrolment step”, the pseudonymization of the document is performed and the document is provided to the Requesting Community.

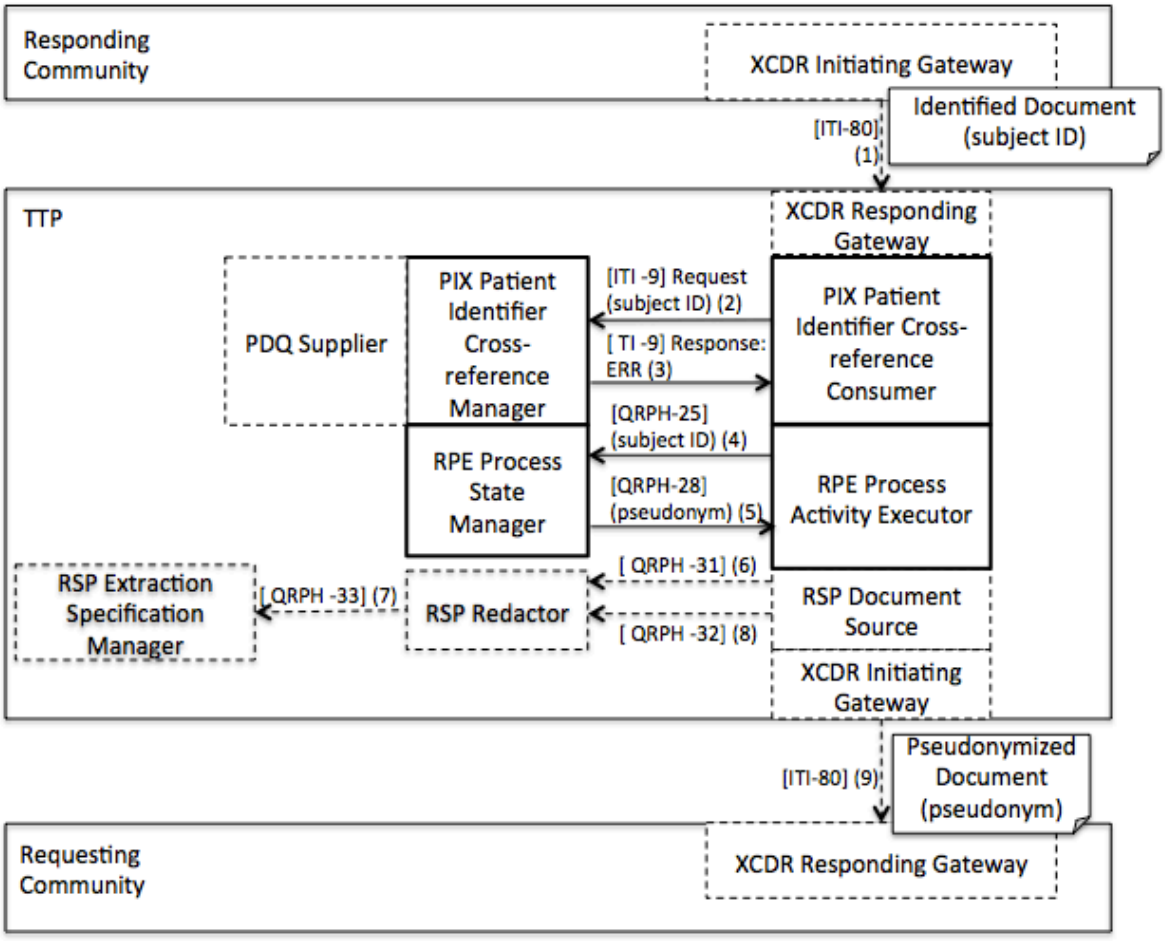


Figure 3.1.4.1-4: Patient enrolment and RPE and PIX profiles (case c)

3.1.5 Cross-Community Semantic Service

Another core activity of the TTP is the management of a semantic service, which provides information about the type of documents storing data of interest in the different communities involved in the study. The IHE profile suitable to be implemented is the DEX (Data Element Exchange) Profile.

3.1.5.1 Data Semantic Management and DEX Profile

Two different use cases are analyzed:

a) The Requesting Community is interested to have access to documents available in the Responding Community;

b) The Requesting Community is interested to have access to data available in the Responding Community.

3.1.5.1.1 Interest on documents: the semantic management (case a)

The Requesting Community needs to know the documents in the Responding Community(ies) storing the data it is looking for. Therefore it should implement the capabilities of a DEX Metadata Consumer which queries with a [QRPH-43] transaction the DEX Metadata Source implemented in the TTP to get a list of clinical data elements matching its needs. After that, the DEX Metadata Consumer chooses from the list, the data element(s) of interest and starts a [QRPH-44] transaction with XCA Document Type Binding Option to ask the DEX Metadata Source for metadata about the clinical data, in particular about the DocumentEntry metadata describing the type of documents storing the clinical data element(s) of interest. This process of retrieval of metadata about data of interest can be performed just once at the beginning of the study when the Requesting Community looks for data matching its needs or it can be performed every time the Requesting Community needs new types of data or in order to chech if metadata have been updated. It depends also on the type of agreeement between the communities: the study protocol might define the exact detail of single data allowed to be queried (in this case the retrieval of metadata can be performed just once) or it might contain more general information for example about the type of documents allowed to be queried (in this case during the study the Requesting Community might need to query metadata about the single data of interest).

The main DocumentEntry metadata returned by the DEX Metadata Source are: classCode, typeCode, formatCode, eventCodeList, homeCommunityID. For example the “hemoglobin data element” may be available in Hematological Laboratory Reports, identified by the following DocumentEntry metadata: classCode “11502-2” (LOINC code for Laboratory Report), typeCode “18723-7” (LOINC code for Hematological laboratory report), eventCodeList “Adult_lab_report” and formatCode “urn:ihe:lab:xd-lab:2008”.

After the DocumentEntry metadata have been retrieved, the XCA Initiating Gateway in the Requesting Community can start a [ITI-38] transaction to the TTP indicating in the homeCommunityID DocumentEntry the identifier of the Responding Community of interest. The TTP then forwards the request with another [ITI-38] transaction to the Responding Community indicated in the homeCommunityID of the request from the Requesting Community. Finally, with the [ITI-39] transaction the documents of interest are retrieved by the Requesting Community.

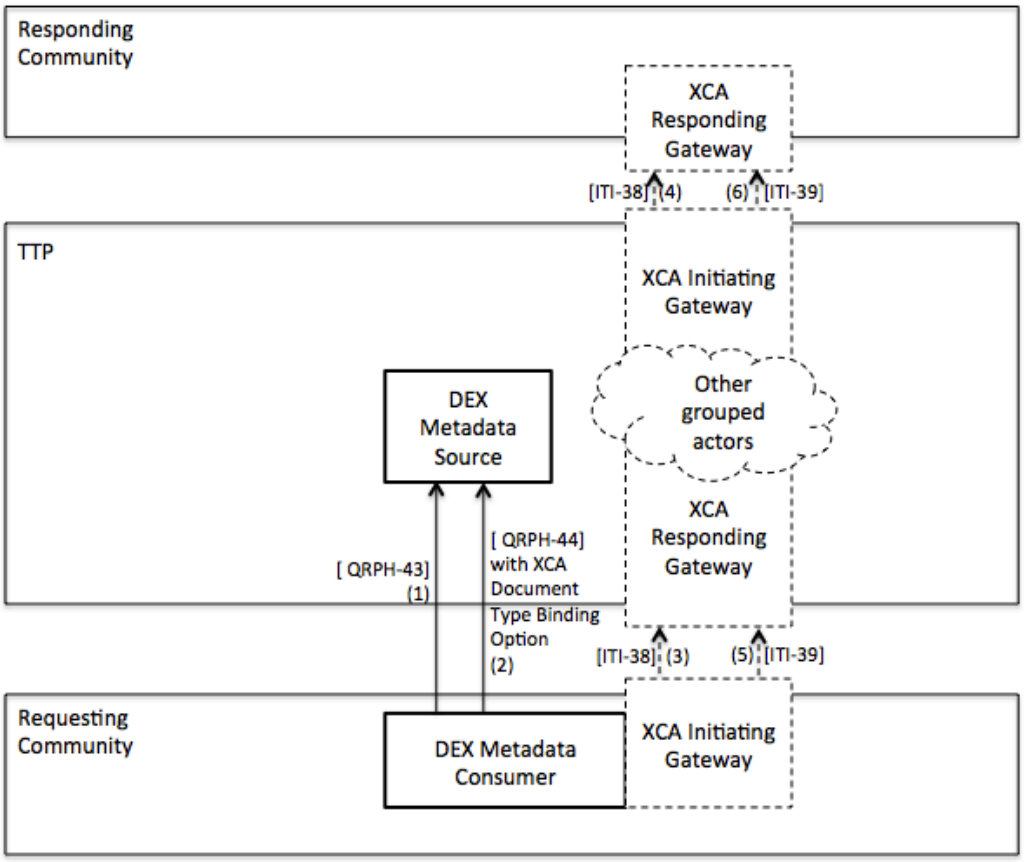


Figure 3.1.5.1.1-1: Semantic management and DEX Profile (case a)

3.1.5.1.2 Interest on data: the semantic management (case b)

The situation here described differs from case a) because the Requesting Community is interested only to retrieve data and not the entire documents storing data. In this case it starts a [PCC-1] transaction to query for clinical data and, it is the TTP in this case that needs to know the DocumentEntry metadata to retrieve the documents containing the clinical data of interest for the Requesting Community. Therefore the TTP has to implement both the capabilities of DEX Metadata Consumer and DEX Metadata Source. In particular, when the TTP receives a request for data from the QED Clinical Data Consumer in the Requesting Community, the DEX Metadata Consumer starts a [QRPH-43] transaction to retrieve the data element(s) matching the needs of the QED Clinical Data Consumer. It then chooses the data element of interest and retrieves with the [QRPH-44] transaction the metadata related to the data elements, in particular the DocumentEntry Metadata. With this information, the grouped XCA Initiating Gateway can starts the process of document retrieval with the [ITI-38] transaction followed by the [ITI-39] transaction. Data is then extracted from documents and provided to the Requesting Community in the [PCC-1] Response.

This process of retrieval of metadata about data of interest can be performed just once at the beginning of the study when the Requesting Community defines data of interest (and the TTP is in charge to look for data matching its needs) or it can be performed every time the Requesting Community needs new types of data or in order to check if metadata have been updated. It depends also on the type of agreeement between the communities: the study protocol might define the exact detail of single data allowed to be queried (in this case the retrieval of data metadata can be performed just once) or it might contain more general information for example about the type of documents allowed to be queried (in this case during the study the TTP might need to query for metadata about the single data of interest).

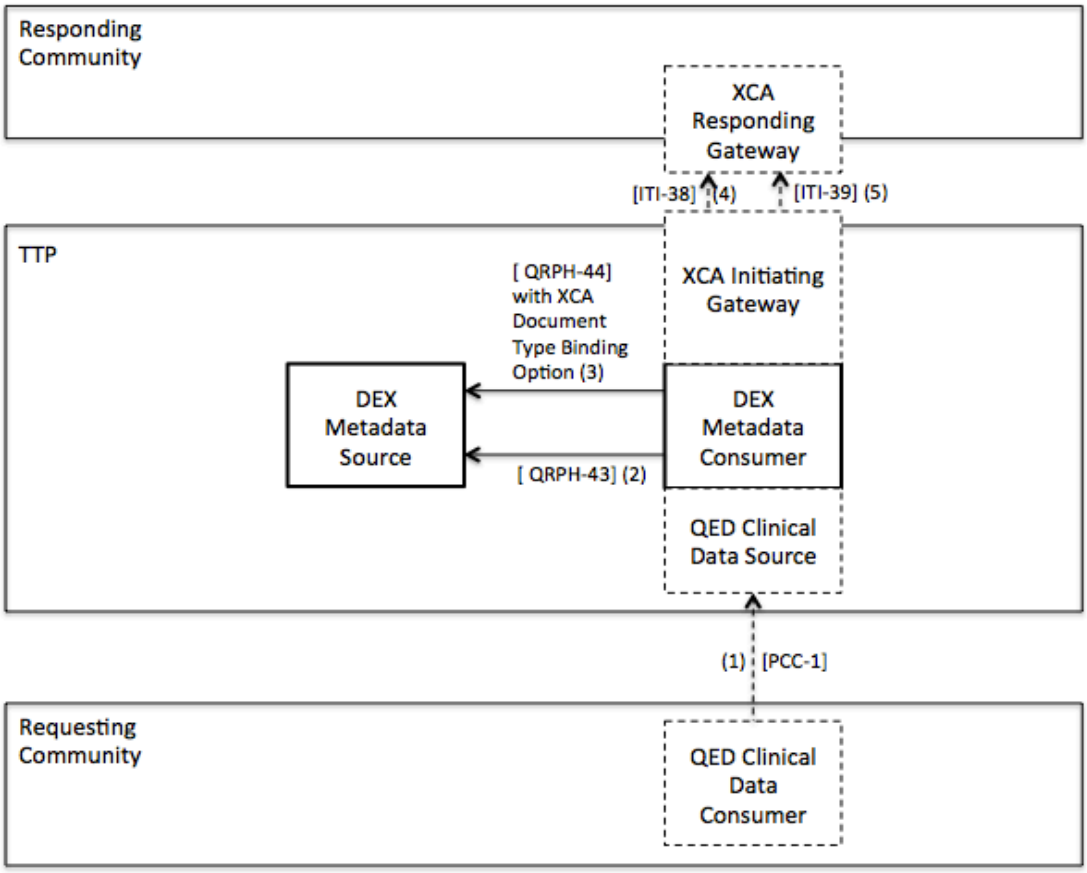


Figure 3.1.5.1.2-1: Semantic management and DEX Profile (case b)

4 Query Definition and Further Architecture Features

The Requesting Community can be interested to define many different types of query. Each query can be defined as described in Figure 4-1.

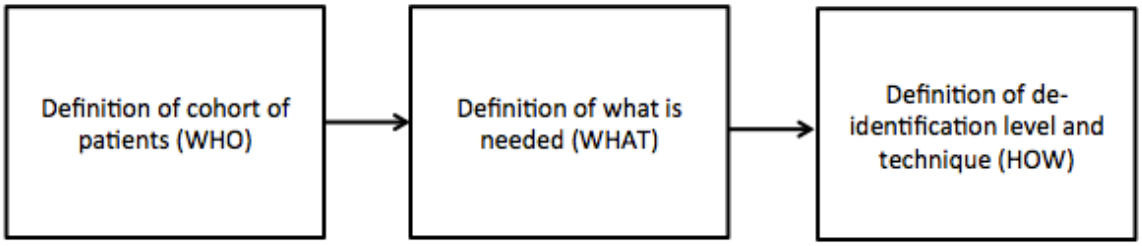


Figure 4-1: Process of query definition

The first step is the definition of the cohort of patients for whom data/documents need to be retrieved.

The cohort of interest can be defined in the following ways:

1) **Specific patients**: the patient identifiers used in the Requesting Community are input parameters of the query. Two subcases can happen depending on privacy needs:
    a. the patient identifier is the same in the Requesting and Responding Community (this is a rare case in secondary data usages) -> the technical solution corresponds to the direct communication between the two communities using the XCA Profile and the involvement of a TTP is not necessary;
    b. the patient identifier used in the Requesting Community is different from that used in the Responding Community -> the technical solution is that presented in the Section 3.
2) There are **not specific patients** for whom data/documents are needed: the query is based on kind of data or documents of interest, or other parameters not related to patients (e.g., the request is about all discharge summaries produced in a specified period in a clinical community)-> the technical solution involves the QED and/or MPQ profiles and is described in Section 4.1.1;
3) A **specific population** matching criteria based on **demographic characteristics**, including for example gender, age, place of birth, place of living, ethnicity (e.g., male people older than 60) -> the technical solution involves the PDQ Profile and it is described inSection 4.1.2;
4) A **specific population** matching criteria based on **clinical characteristics** (e.g., diabetic people) -> the technical solution involves many profiles (including QED and MPQ) and is described inSection 4.1.3.

The second step is the definition of what is needed:

1) **Specific types of documents** -> the technical solution is that presented in Section 3.1.2.1 (usage of XCA and the types of documents are defined by the DocumentEntry metadata);
2) **Specific data** -> the technical solution is that presented in Section 3.1.2.3 (usage of QED and XCA).

The third (logical) step is the definition of how data/documents are provided in relation to the type of de-identification technique used (if any). No IHE profiles have been provided until now which allow to specify the type of de-identification technique to be used, however this is not a big issue since it is usually agreed at the beginning of the study (so temporally it is the first step of the query definition) and should be defined using the CRPC Profile (as described in Section 3.1.1.1).

Data can be provided as:

1) **Identified** data/documents: the patient identifier used in both the Requesting and Responding Community is the same, so no de-identification technique is needed (this is a rare case in secondary data usages) -> the technical solution corresponds to the direct communication between the two communities using the XCA Profile and the involvement of a TTP is not necessary;
2) **Pseudonymized** data/documents: the patient identifier used in the Requesting Community (the pseudonym) is different from that used in the Responding Community -> the technical solution is that presented in Sections 3.1.3.1 and3.1.4.1;
3) **Anonymized** data/documents: data/documents are provided to the Responding Community without any patient identifier (or in case with an "anonymous identifier") -> the technical solution is that presented in Section 3.1.3.1;
4) **Aggregate data**: data are not provided related to specific patients but as aggregate data -> the technical solution is presented in Section 4.2 and involves the ADX Profile.

The process of query definition described above is not applicable in the case of the **"push" solution** presented in Section 3.1.2.2: in this case the criteria about the patient/population of interest, data/documents of interest, the de-identification technique to use and when documents are sent to the TTP, are defined the beginning of the study in the study protocol. The Community providing data should implement upstream a filter to select documents and patients according to the pre-defined criteria. Otherwise it is the TTP that has to provide some functionality in order to be able to filter the documents of interest, e.g., in order to identify people of interest for the Requesting Community, it could implement PDQ capabilities or it could extract directly from document metadata or from the document content itself some useful information allowing to select only people of interest or to select document matching other types of criteria, as temporal criteria.

This chapter will show the IHE standard technical solutions that can be adopted to satisfy the different types of requests above indicated. Specifically, it focuses on the functionalities to be

provided by the TTP, the Responding, and the Requesting Community, that have not been already presented in Section 3: they are usually secondary functionalities, even if in specific use cases they can be required.

4.1 Definition of cohort of patients

The technical solution for case 1) (from the list at page 48) has been already defined in the previous chapter. This chapter shows the technical solutions for cases 2, 3 and 4.

4.1.1 Unspecified patients

This section focuses on case 2) in the list indicated at page 48. An exemplifying use-case about the retrieval of documents for unspecified patients is a Regional health authority that wants to have access for governance purposes to all the discharge summaries produced in a specific period in a HIE system.

This section focuses on use cases when the Requesting Community does not specify any patient's identifier, since the request is for all patients belonging to the Responding Community and not known a priory by the Requesting Community because any specific patient has been enrolled in the study (e.g., an epidemiology organization which would like to have access to anonymous discharge summaries produced in a HIE system and it is not interested on the patients' they belong to). On the contrary, if the Requesting Community does not want to specify any patient, but the patients' enrolment has been performed, so it knows its own patients belonging also to the Requesting Community, the technical solution is just that presented in Section 3.1.2.1, where the transaction is repeated for each patient belonging to the Requesting Community and enrolled in the study.

The XCA Profile does not allow to perform a [ITI-38] transaction without specifying the patient identifier, on the contrary the MPQ Profile has been actually intended to perform this kind of queries and it seems the most suitable profile to be applied. However, it requires the Requesting Community, the TTP, and the Responding Community, to be organized as Document Sharing environments and this may be a very strong assumption mostly for Secondary Data Usage Communities. The technical solution is shown in Figure 4.1.1-1. If in the near future a change proposal to XCA will allow to not specifying the patient identifier (some discussions are ongoing about it), the technical solution will be the same of the query for documents related to specific patients and shown in Figure 4.1.2-1.

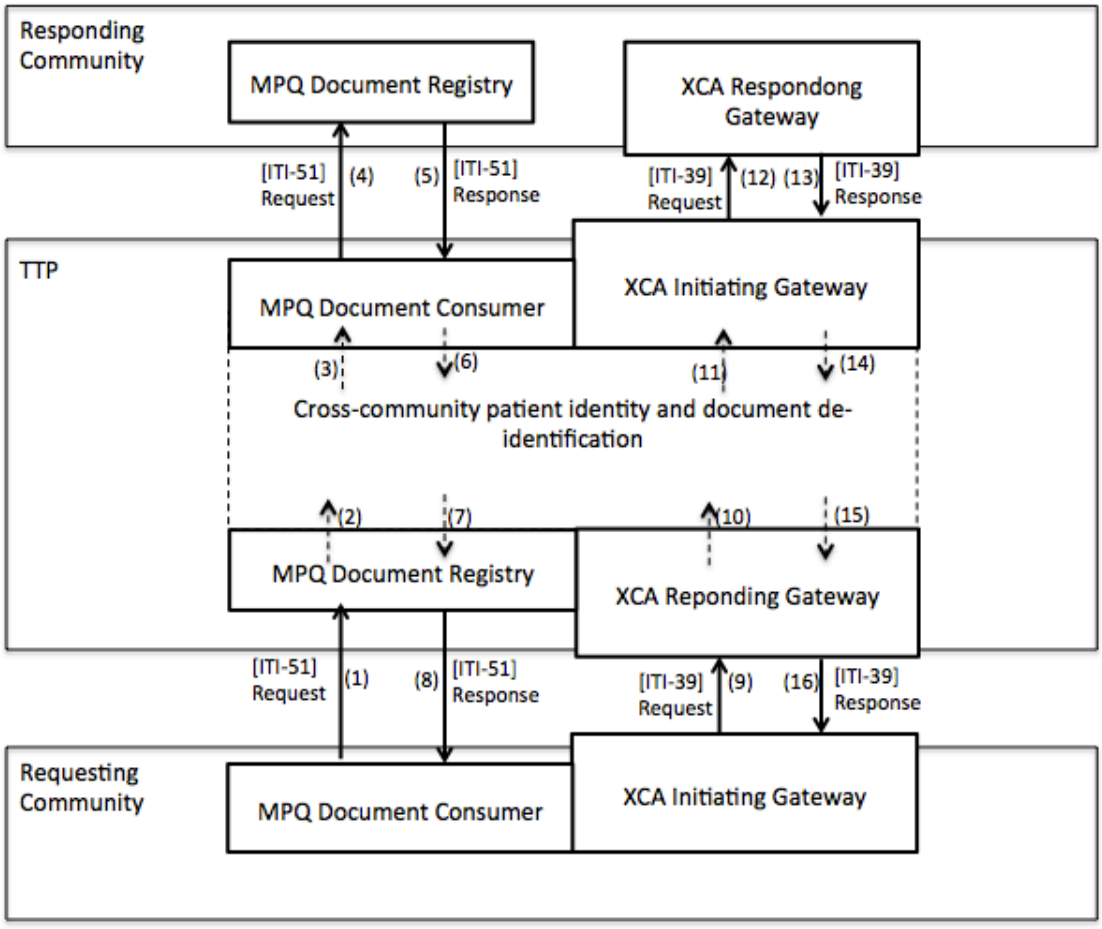


Figure 4.1.1-1: Document retrieval for unspecified patients

An exemplifying use-case about the retrieval of data, instead of documents, for unspecified patients is a Regional health authority that wants to have access for governance purposes to all diagnosis at hospital discharge performed in a specific period in a HIE system.

The technical solution in case of a query about data retrieval for unspecified patients is shown in Figure 4.1.1-2. The QED Profile allows to not specify the patient identifiers as input of the query (Multi-Patient Query Option), so it can be used to retrieve clinical data also when no specific patients are of interest. The TTP is then in charge to ask to the Responding Community for documents storing data of interest, so both the TTP and the Responding Community should implement the MPQ Profile functionalities in order to retrieve the registry entities with the [ITI-51] transaction and XDS Profile functionalities in order to retrieve the documents of interest with the [ITI-43] transaction. Data is then extracted from documents and de-identified before being provided to the Requesting Community in the [PCC-1] Response.

Another solution that does not require MPQ and XDS functionalities is that the TTP has PDQ capabilities and knows the identifiers for patients belonging to the Responding Community: in this case, the XCA Profile can be used specifying the patients' identifiers as input query parameter (see Figure 4.1.2-1).

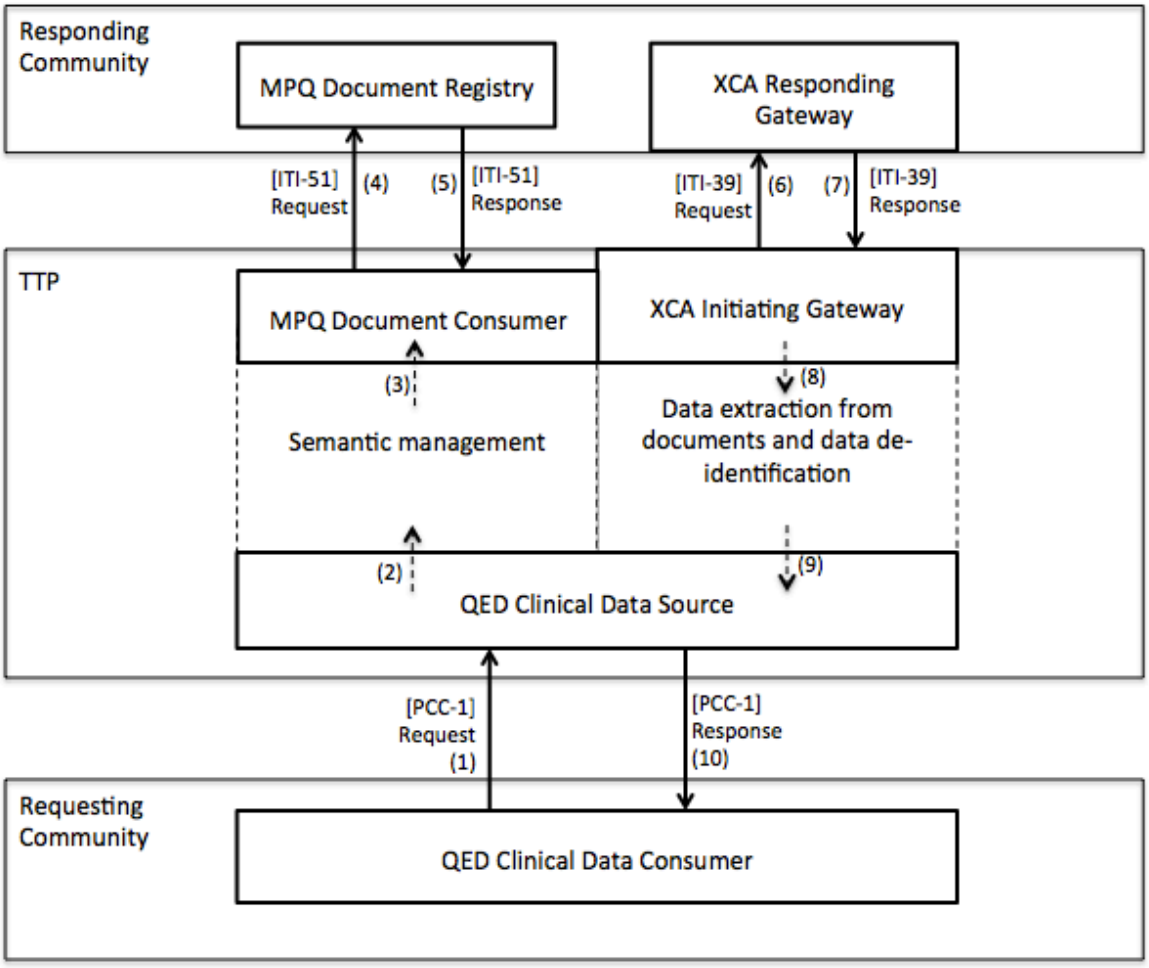


Figure 4.1.1-2: Data retrieval for unspecified patients

4.1.2 Cohort definition according to demographics characteristics

An exemplifying use-case about the retrieval of documents for patients matching specific demographic criteria is a Regional health authority that wants to have access for governance purposes to discharge summaries related to male patients older than 65 in a HIE system.

The XCA, QED and MPQ profiles, which are the three IHE profiles that can be implemented in order to retrieve documents/data of interest, do not allow specifying demographic patient's characteristics as query parameters (in the PCC-1 transaction of the QED Profile the

patientAdministrativeGender and patientBirthTime demographic query parameters are meant to be used only to validate the patientID and, if the patient identifier is not valued, they shall not be used). Therefore, first, the Requesting Community has to identify patients matching the demographic criteria and, secondly, to ask for data/documents related to these patients.

The PDQ Profile is the suitable profile to be used since it is aimed to provide the identifiers of patients matching demographic research criteria: the Requesting Community has to implement the PDQ Consumer, the PDQ Supplier can be played either by the TTP or by the Requesting Community itself.

If the PDQ Supplier is played by the TTP, this one might have obtained the demographic information in three different situations:

1. During the enrolment phase as described in case a) and b) at page 37;
2. During previous documents retrieval as described in case c) at page 40;
3. Other internal mechanisms.

If the interest is about document retrieval (e.g., all discharge summaries for male patients), the technical solution is depicted in Figure 4.1.2-1, if the interest is on data retrieval (e.g., diagnosis of diabetes for females), the technical solution is represented in Figure 4.1.2-2.

In this specific solution, the PDQ Supplier has to provide information only about patients belonging to the specific Requesting Community. On the contrary, the PDQ Profile states that if in the [ITI-21] Request the domain is not specified, the PDQ Supplier shall provide information about all domains involved. This requirement implies that the TTP shall implement different PDQ Supplier actors, one for each community involved in the study.

Another technical solution is that, first, the Requesting Community asks for documents not specifying any demographic characteristics (as described above in Section 4.1.1) and, then, it extracts demographic information from metadata or data stored within documents: in this way it can filter the documents related to patient of interest. However, this solution is slower and not optimal, since it requires consuming lots of resources to retrieve documents for all patients and then to detect only data/documents for patients really matching the research criteria.

Interest on documents

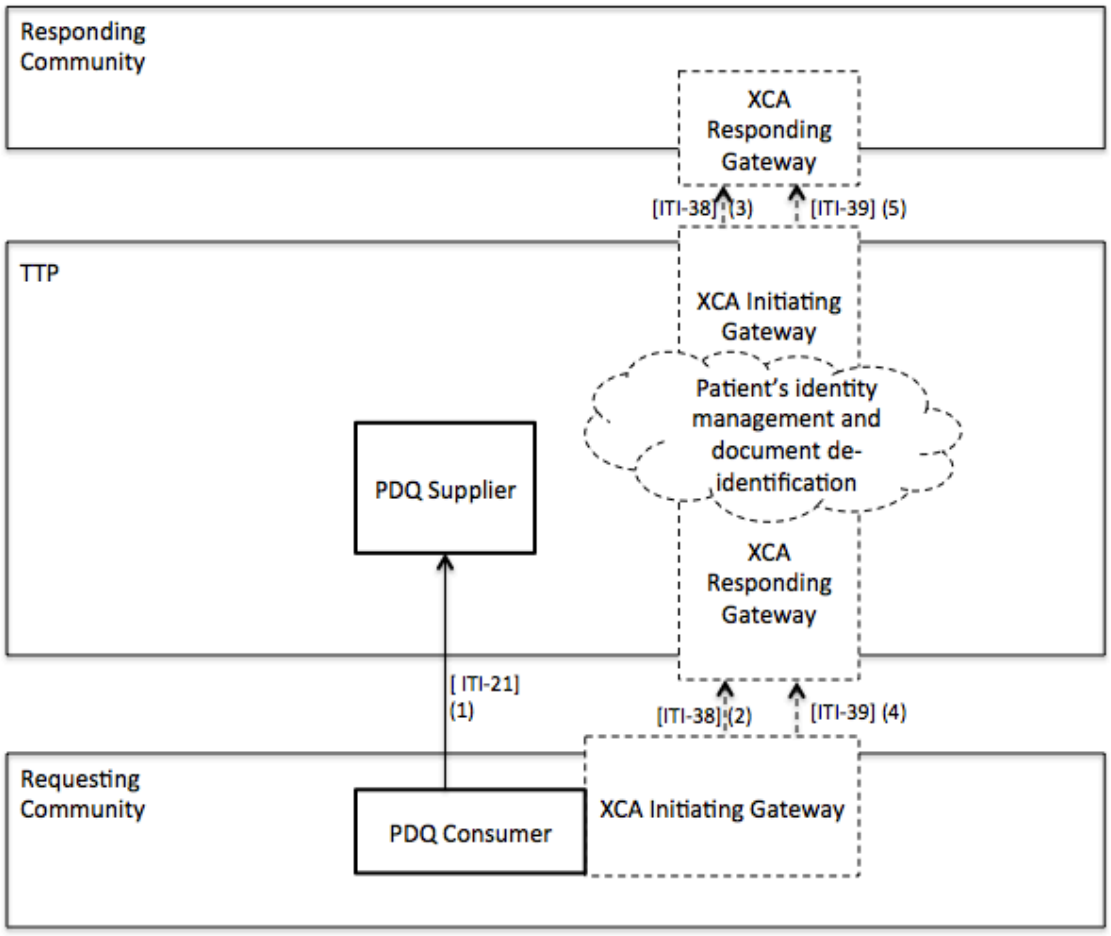


Figure 4.1.2-1: Documents retrieval for a cohort defined on demographic characteristics

Interest on data

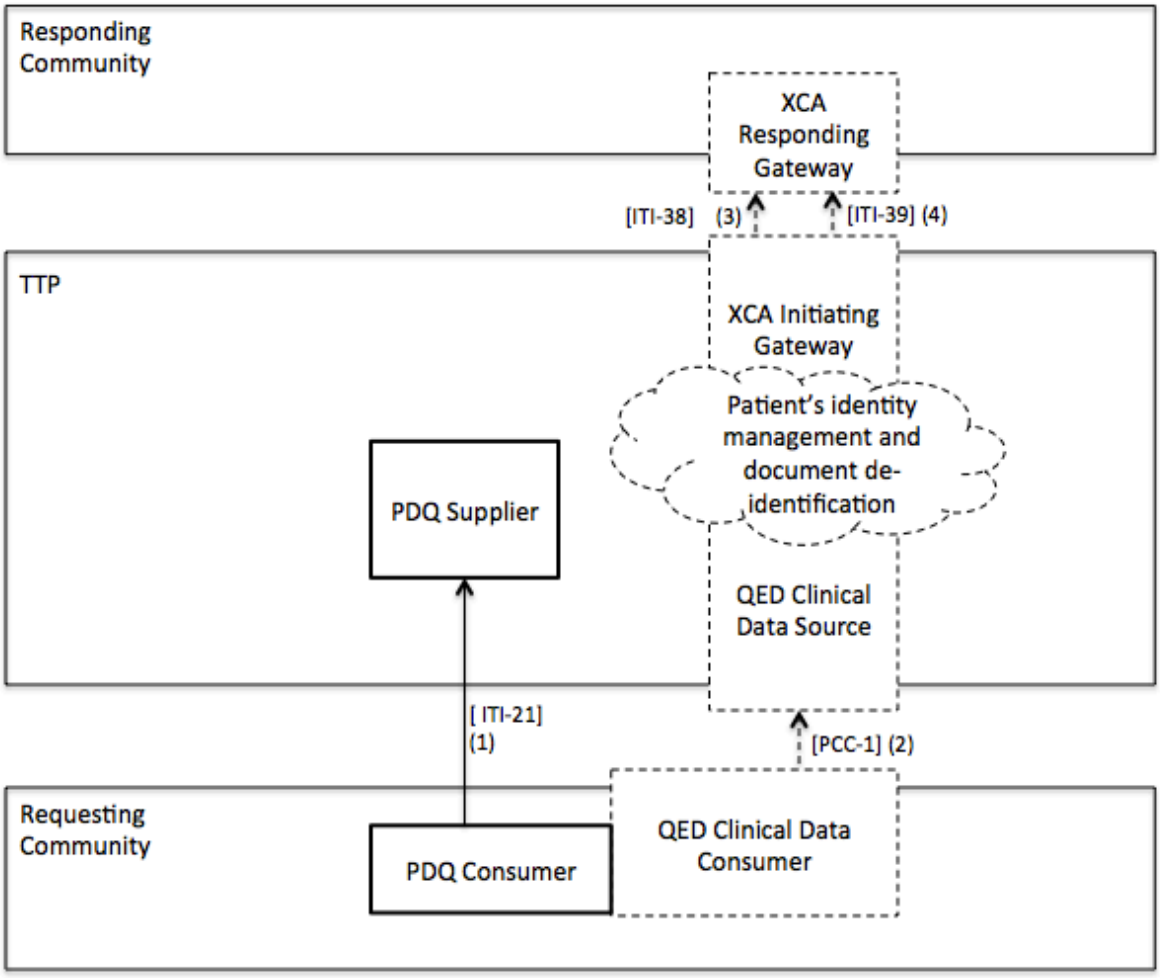


Figure 4.1.2-2: Data retrieval for a cohort defined on demographic characteristics

4.1.3 Cohort definition according to clinical characteristics

An exemplifying use-case about the retrieval of documents for patients matching specific clinical criteria is a research organization that wants to perform a study on diabetes and would like to have access to discharge summaries related to diabetic people and produced in a HIE system.

The definition of a cohort according to clinical characteristics can be performed on the basis of specific clinical data that can be retrieved by the Requesting Community from the Responding Community. For example, if the cohort of interest is composed of diabetic people, the clinical data useful to define this cohort is a diagnosis of diabetes. Therefore in order to retrieve discharge summaries related to diabetic people, first of all, the cohort of diabetic people has to be identified on the basis of the "diagnosis of diabetes" clinical data (retrieved with a [PCC-1]

transaction with the Multi-Patient Query Option, as indicated in Section 3.1.2.3.1 and then the retrieval of discharge summaries for these patients is performed (using the XCA Profile, as indicated in Section 3.1.2.1). If the TTP has also PDQ functionalities, the identification of the cohort of interest can be split in two steps in order to improve the computation time of the [PCC-1] transaction: first, a [ITI-21] transaction can be sent by the PDQ Consumer in the Requesting Community in order to identify patients matching some demographic criteria and then, only for these patients the, diagnosis of diabetes is looked for in the [PCC-1] transaction.

The technical solution is the combination of profiles and processes defined in the previous sections and it is represented in Figure 4.1.3-1 in case of retrieval of documents and in Figure 4.1.3-2 in case of retrieval of data (specifically they consider also the first optional step of the [ITI-21] transaction, however this step can be omitted). A [PCC-1] transaction is initiated by the QED Clinical Data Consumer in order to retrieve the clinical data useful to identify the cohort of interest, followed by the [ITI-38] and [ITI-39] transactions between the XCA Initiating Gateway in the TTP and the XCA Responding Gateway in the Responding Community to retrieve documents useful to extract the “diagnosis” clinical data and provide the final data to the Requesting Community in the [PCC-1] Response. With this information, the Requesting Community defines the final cohort of interest: after that, the usual process of documents retrieval can be performed with the [ITI-38] and [ITI-39] transactions (Figure 4.1.3-1) or the usual process of data retrieval with the [PCC-1] transaction (Figure 4.1.3-2).

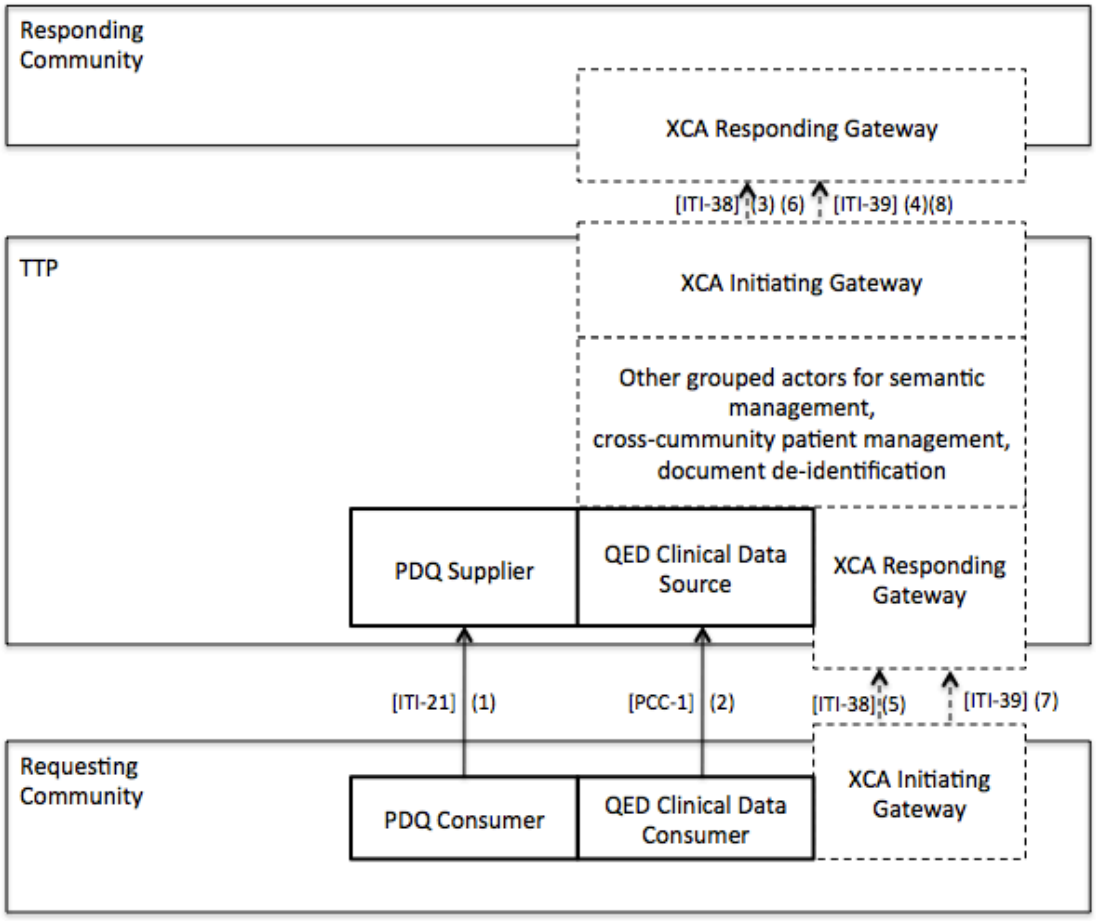


Figure 4.1.3-1: Documents retrieval for a cohort defined on clinical characteristics

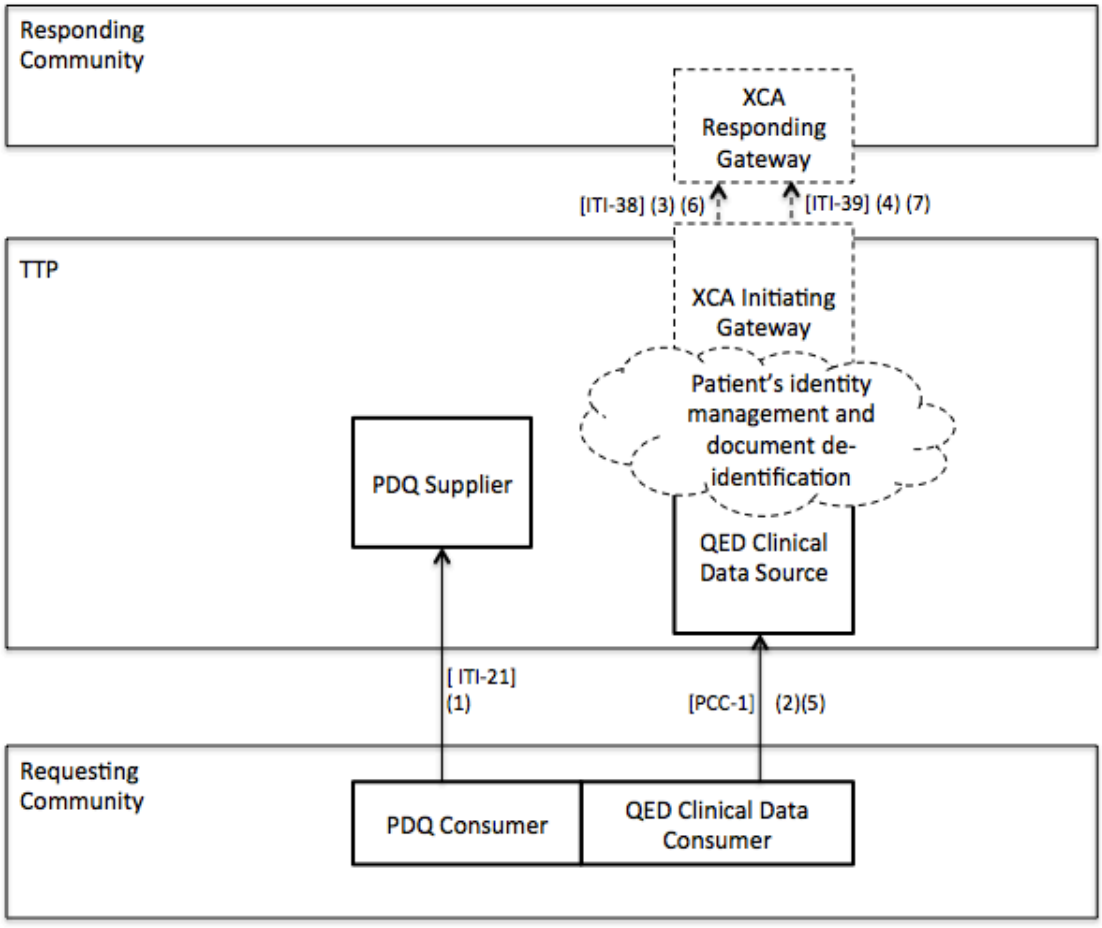


Figure 4.1.3-2: Data retrieval for a cohort defined on clinical characteristics

4.2 De-identification technique: aggregate data

In this section the provision to the Requesting Community of data as an aggregate report is presented. Aggregation of data is usually performed when a very strong and secure de-identification technique is needed and/or when there is no need for the Requesting Community to have access to patient level data and/or when the interest of the Requesting Community is only to perform measurements about the Responding Community, e.g for quality purposes.

The aggregation of data can be performed both by the TTP or directly by the Responding Community. However, the involvement of the TTP is be very useful when also a cross-community semantic service is needed in order to identify documents in the Responding Community containing data of interest for the Requesting Community. Moreover, the involvement of the TTP is necessary when it is involved in the study for other activities (as patients' identity management and study management): in this case it is more straghtforward and easier if all the activities about documents retrieval, data extraction and data aggregation are also

performed by the TTP. This is in particular the case if the Requesting Community is interested to have access as an aggregate report to data related to specific patients belonging to both the two communities, so when a linkage at the beginning of the study has to be performed by the TTP in order to identify correspondent patient's identifiers in the two communities (according to the solution presented inSection 3.14.1, case b). So, since the patient's management is performed by the TTP, it should also be in charge to perform queries for documents related to the patients enrolled in the study and to produce the final report.

Moreover, a TTP may be very useful in order to guarantee the quality of the final aggregate report: the involvement of a "third" party, which is neutral in the study and uses pre-defined and transparent queries (agreed at the beginning of the study) to retrieve dcouments of interest from the Responding Community, assures that the finale aggregate report is highly reliable and complete. Finally, the involvement od the TTP may reduce the burden of the Responding Community about the creation of aggregate reports.

The technical solutions presented below are built on the hypothesis that is the TTP in charge to create the aggregate report. If this functionality is directly provided by the Responding Community, the solution is similar since the actors played in the other case by the TTP, are in this case played by the Responding Community.

The QRPH domain developed two profiles to define how to create, share and consume aggregate reports in a standard way. The first profile is Aggregate Data Exchange (ADX), the second one is Quality Metrics Execution – Early Hearing (QME-EH).

ADX enables interoperable public health reporting of aggregate health data. ADX will typically be used to represent routinely reported aggregate data such as the numerators and denominators, which can be used in the construction of public health indicators. ADX defines a Content Data Structure Creator Actor that creates two message structures that enable an implementing jurisdiction to formally define the aggregate health data to be exchanged:

- ADX profiles the SDMX v2.1 Data Structure Definition (DSD) specification
- ADX normatively describes how a DSD file is transformed to develop an XML schema definition (XSD) file

ADX Content Creator and ADX Content Consumer Actors use the DSD and XSD to construct and exchange ADX/XML messages containing aggregate health data in their jurisdiction.

The QME-EH Content Profile specifies how to create and consume standard electronic patient-level and aggregate-level quality reports for the Newborn Hearing Screening (CMS31v4) electronic clinical quality measure (eCQM). It also specifies how to reuse data from a standard summary of care document generated by an EHR to create a patient-level quality report. Additionally it specifies how to create an aggregate-level quality report for the Newborn Hearing Screening quality measure from multiple patient-level quality reports.

The technical solution proposed here implements the ADX Profile, since this can be applied for general purposes. On the contrary the QME-EH Profile defines the process of creation of quality reports for a specific purpose: the Newborn Hearing Screening electronic clinical quality measure, so it is suitable when this is the specific purpose of the secondary data usage.

ADX requires an exact specifications of data to be aggregated and, in general, about the content of the aggregate report. Specifically, it defines the ADX Content Data Structure Creator which is in charge to profile the DSD specification and how it is tranformed to develop the XSD file. Therefore, at the beginning of the study the two communities and/or the TTP have to agree exactly on how the report has to be built and one of the parties (usually the TTP or the Requesting Community) is in charge to play the role of the ADX Content Data Structure Creator (in Figure 4.2-1, Figure 4.2-2, Figure 4.2-3, which show three different scenarios according to the type of population of interest as described below, for example the ADX Content Data Structure Creator is played by the TTP). This actor defines the DSD specfications and share this content to the ADX Content Creator played by the TTP and the ADX Content Consumer played by the Requesting Community.

After this first stage about the definition of the content module specifications, the TTP has to identify which are the data element available in the Responding Community that allow to create the aggregate report, so both the data that are actually the outcome of the measurement and data which serve to stratify data and to create the classes in which data are grouped (e.g., spatial data about the health facility pruducing the outcome, temporal data about the time of data creation, population data about the patient's age). After that, the TTP has to identify the clinical documents in the Responding Community containing the data of interest. The DEX Profile is applied for this purpose as already presented in Section 3.1.5.1.2, where both the DEX Metadata Consumer and DEX Metadata Source are played by the TTP (Figures 4.2-1,4.2-2 and 4.2-3). The [QRPH-43] and [QRPH-44] transactions are usually sent by the Metadata Source just once at the beginning of the study in order to retreive the information of metadata of interest (sometimes they are sent every time a new report has to be created in order to check if metadata in the meantime have been updated). However in Figures 4.2-1,4.2-2 and 4.2-3 they are indicated as the first and second transaction since conceptually these are the first transactions to be performed in order to allow documents retrieval.

ADX is a content profile and it does not define transactions to initiate a process of creation of an aggregate report. Therefore at the beginning of the study, the time schedule for the creation and sharing of the aggregare report has to be defined in the study protocol.

Figures 4.2-1,4.2-2 and 4.2-3 present three technical solutions about the process of data aggregation, they differ basically on the type of population for which data are needed and in how the population is determined.

In case a) (Figure 4.2-1) the eligibility criteria about the population of interest is defined in the study protocol, but no specific patients are enrolled at the beginning of the study. An examplificative use case is a Requesting Community interested to retrieve a report as aggregate data showing the prevalence of diabetes in all the population older than 65 living in the Responding Community, stratified by year of age and gender. In this case, the TTP plays the role

of the MPQ Document Consumer, which asks to the MPQ Document Registry played by the Responding Community about documents matching specific criteria (e.g., based on the patient's date of birth), as previously presented in Section 4.1.1. After that, a [ITI-39] Request is sent by the XCA Inititating Gateway in the TTP to retrieve documents conataining data of interest in the Responding Community. After documents are retrieved, data are extracted and the ADX Content Creator played by the TTP creates the aggregate report according to specifications previously defined and shared by the ADX Content Data Structure Creator. The aggregate report is shared with the ADX Content Consumer played by the Requesting Community.

In case b) (Figure 4.2-2) the population of interest is based on demographic criteria as in case a), but in this specific case the TTP has also PDQ funcionalities. Demographic information may have been collected in different ways, for example if an enrolment phase has been performed in previous studies involving the two communities or at the beginning of the specific study (case a and b) at page 37), or from previous documents retrieval (case c) at page 40 or from other local mechanisms. The PDQ Supplier informs in the [ITI-21] Response the PDQ Consumer (both the two actors are played by two different systems in the TTP) about the identifier of patients in the Responding Community matching the eligibility criteria. Then a [ITI-39] Request is sent by the XCA Inititating Gateway in the TTP to retrieve documents conataining data of interest in the Responding Community for patients previously identified by the PDQ Supplier. After the documents are retrieved, data are extracted and the ADX Content Creator played by the TTP creates the aggregate report according to specifications previously defined and shared by the ADX Content Data Structure Creator. The aggregate report is shared with the ADX Content Consumer played by the Requesting Community.

In case c) (Figure 4.2-3) the population of interest is composed by specific patients enrolled in the study (or in previous studies involving the two communities) and for those a linkage between the patient's identifiers in the two communities has already been performed (as described in Section 3.1.4.1 case b). An exemplificative use case is a Research Community that would like to have access to data regarding some patients enrolled in previous clinical trials, but in the current study the Clinical Community make available to the Research Community further clinical data (e.g., about the patients' immunization status) only as aggregate data. The information of patients enrolled in the study is stored and managed by the PIX Patient Identifier Cross-reference Manager played by the TTP. For these patients, as in the previous cases, a [ITI-39] Request is sent by the XCA Initiating Gateway in the TTP to retrieve the documents containing the data of interest in the Responding Community. After documents are retrieved, data are extracted and the ADX Content Creator played by the TTP creates the aggregate report according to specifications previously defined and shared by the ADX Content Data Structure Creator. The aggregate report is shared with the ADX Content Consumer played by the Requesting Community.

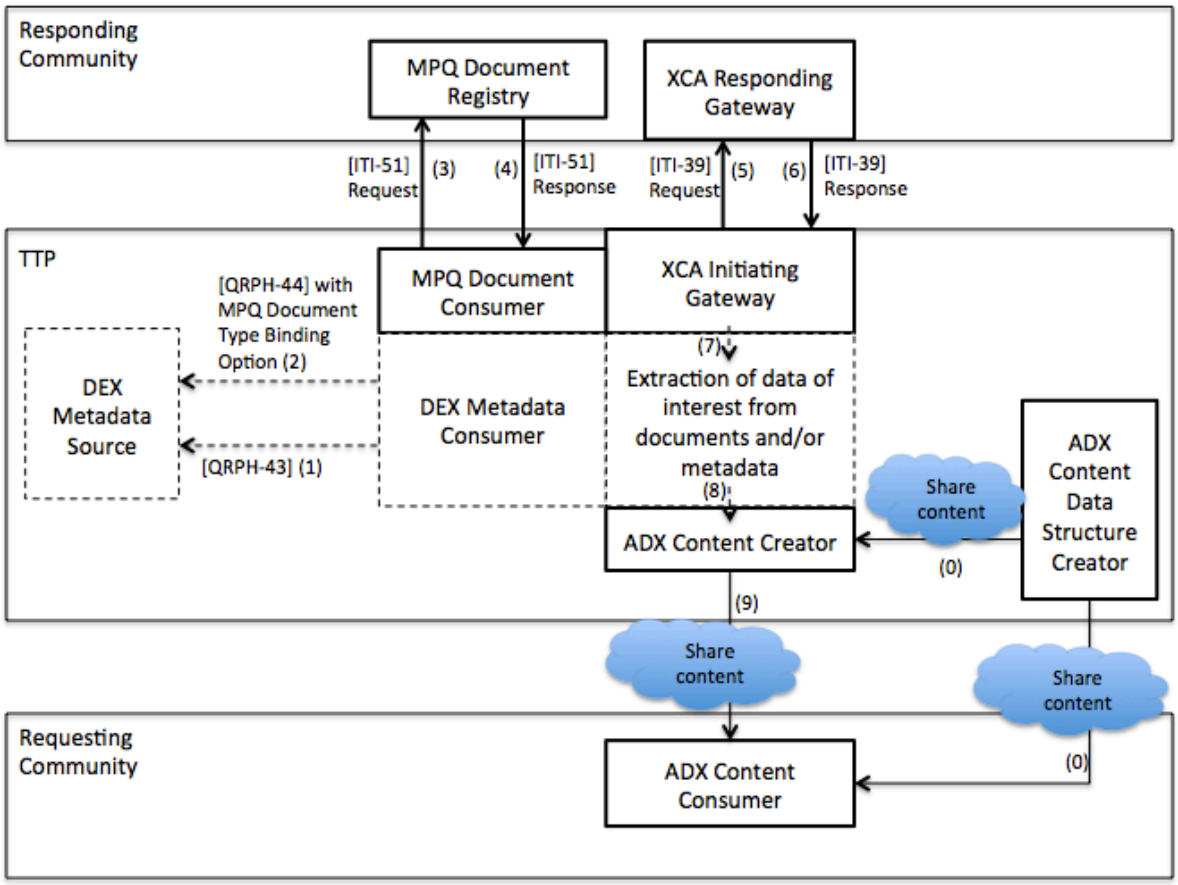


Figure 4.2-1: Data provided as aggregate data (case a)

1635

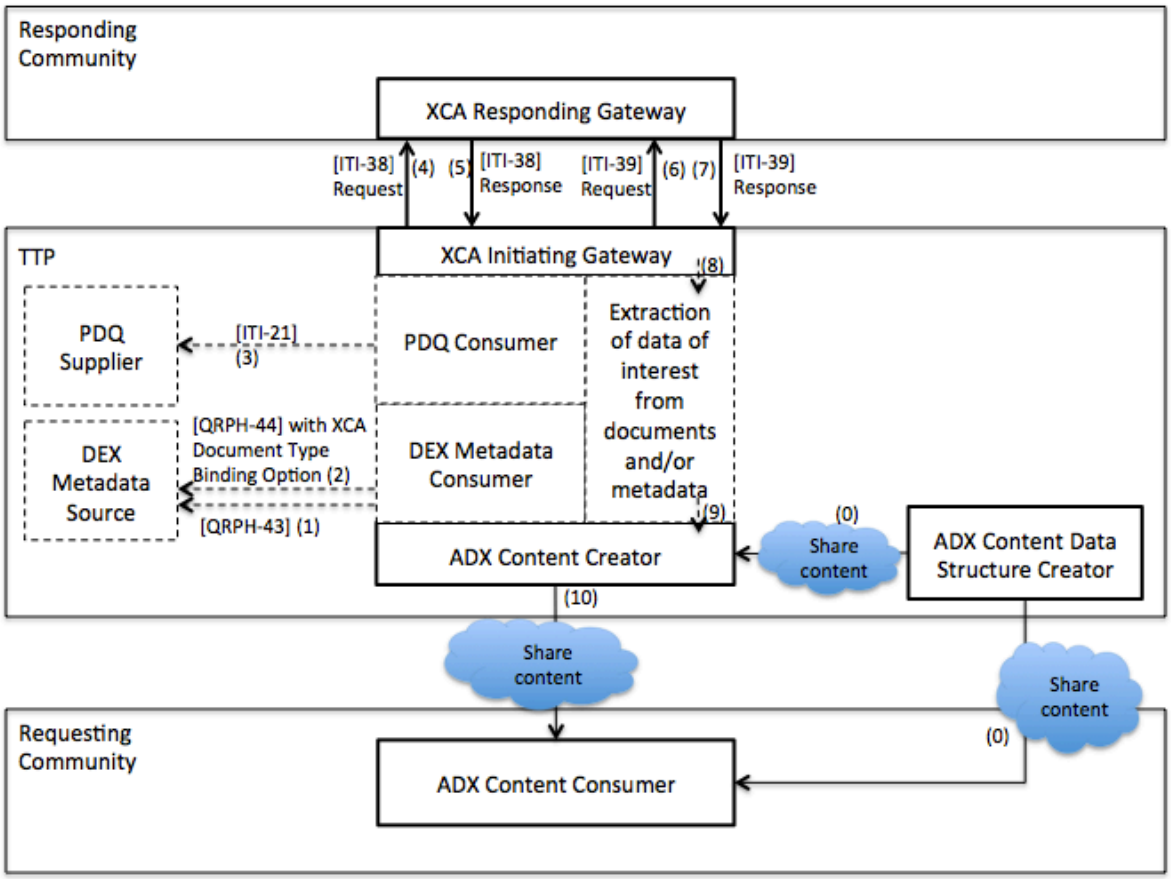
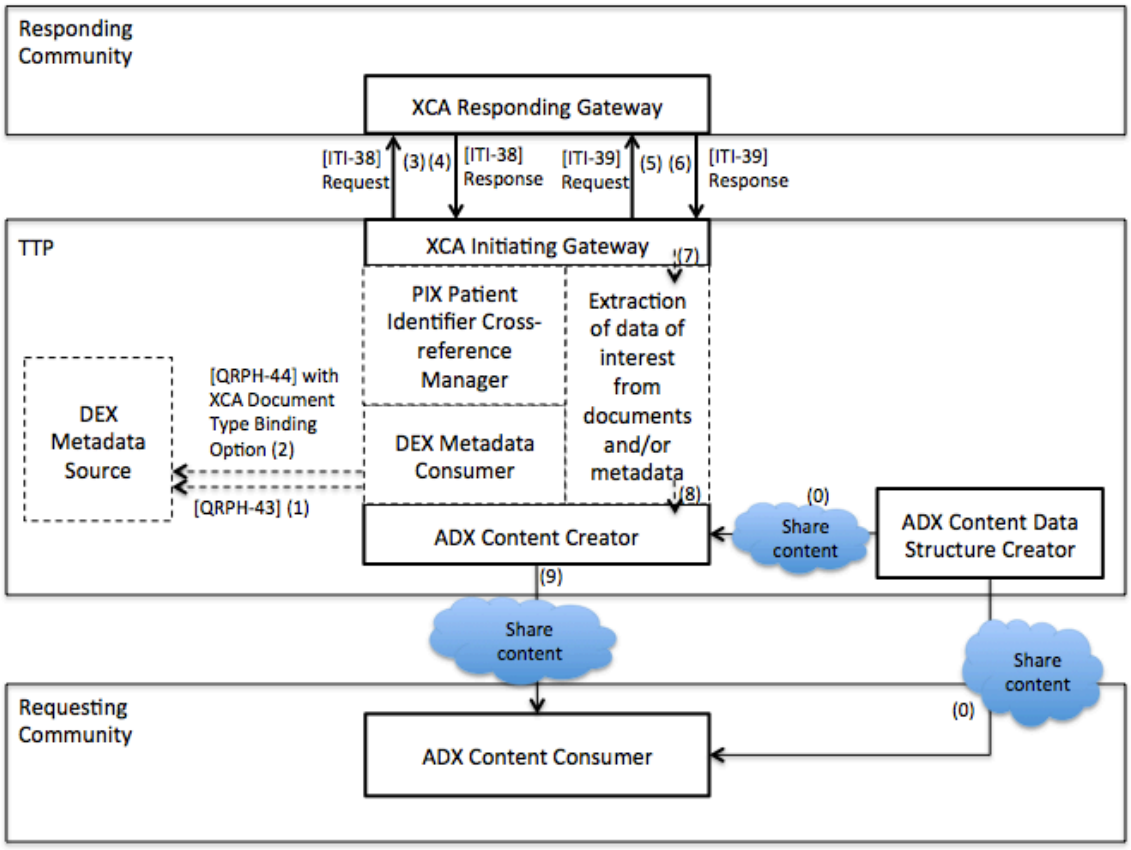


Figure 4.2-2: Data provided as aggregate data (case b)



1640 **Figure 4.2-3: Data provided as aggregate data (case c)**

5 Technical Solutions For Exemplifying Use Cases

5.1 Complete IHE architecture

Combining all the functionalities presented in the previous chapters, the complete standard architecture solution becomes that represented in Figure 5.1-1. Not all the actors and transactions represented in figure should then be implemented in real scenarios, it depends on the specific needs of each real situation.

The different colours highlight the different main functionality of each actor:

- Study Management functionality (green);
- Patient Identity management (red);
- Data/document de-identification (orange);
- Data/document retrieval and provision (light blue);
- Semantic management (blue).

In the following sections (5.2 and 5.3) two use-cases and the proposed IHE solutions to be implemented for the specific scenarios are presented.

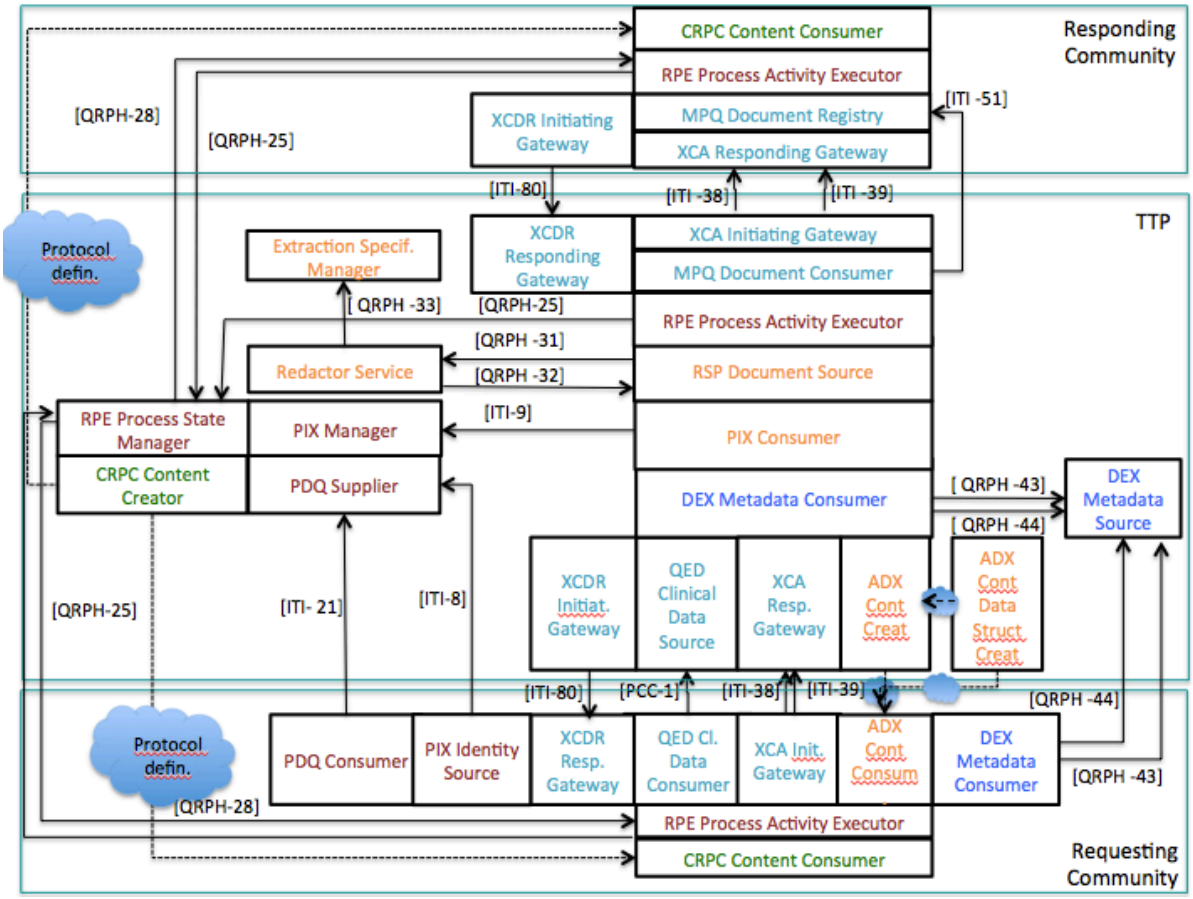


Figure 5.1-1: Complete standard architecture

5.2 Use case 1: epidemiological study

The first use case is about a researcher that would like to calculate the prevalence of diabetes type 1 within his/her community. In the current state, it is not a trivial task since an observational study is usually performed and cumbersome procedures have to be established: they take a long time, include a long follow-up, and involve a lot of people and resources. For example, in order to evaluate the prevalence of gestational diabetes mellitus in relation to race and socioeconomic status in the Region where he lives, a cross-sectional study is usually performed: lots of women are enrolled in the study, many hospitals may be involved in order to reach a sufficient sample size and women should be screened for all the pregnancy period. In the HIE system established in the Region different kinds of documents are produced: Discharge Summaries, ER Referrals, ePrescriptions, eReferrals, Laboratory Reports, Pathological Anatomy Reports, Vaccination reports, which may contain the diagnosis information, which would allow the researcher to calculate the prevalence of diabetes.

The IHE standard technical solution that would allow the researcher to have access to data to answer his research questions is represented in Figure 5.2-2 and Figure 5.2-3. Specifically, the Figure 5.2-2 highlights the IHE actors to be implemented to solve this specific use-case, the Figure 5.2-3 shows only the specific actors involved in the process of query for data retrieval and the order of transactions flow.

First, the Research Organization has to ask to the administrative authority of the HIE system to have access (in respect of legal and privacy issues) to the clinical data needed to answer his research questions. A protocol has to be defined and agreed by the parties (and approved by Data Protection Authority and/or Ethics Committee if needed) and created according to the CRPC Profile by the TTP where it is stored. In this specific example, the protocol defines that only anonymous data have to be provided as anonymous data.

After this preliminary step, the researcher can start to perform queries to the TTP in order to retrieve information about the patients with a diagnosis of diabetes type 1. The conceptual schema of the query is indicated in Figure 5.2-1.

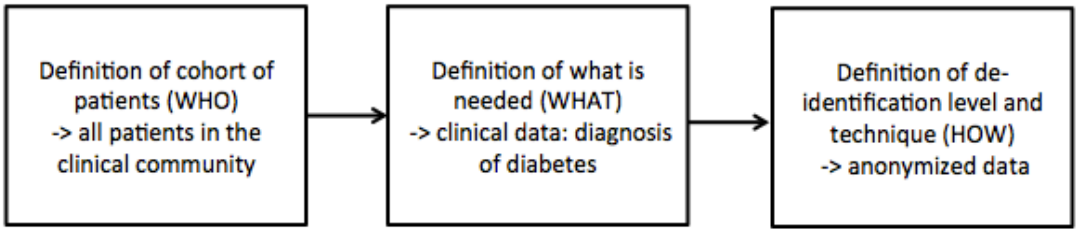


Figure 5.2-1: Query definition for use case 1

From a technical point of view, the query performed by the researcher to retrieve the data of interest is implemented as a [PCC-1] transaction where both the “Problems and Allergies Option” and the “Multi-Patient Query Option” are claimed. Specifically, in the [PCC-1] Request the element <CareProvisionCode> is valued as “PROBLIST” and the “250.01” value (ICD9CM code for diabetes type 1) is indicated as the specific problem looked for. The researcher might also specify in the <clinicalStatementTimePeriod> the period of interest related to the time the diagnosis has been performed.

In order to provide the clinical data of interest, the DEX Metadata Consumer (grouped with the QED Clinical Data Source in the TTP) asks with the [QRPH-43] transaction to the DEX Metadata Source about the list of “data elements” available in the Clinical Community and corresponding to the PROBLIST data element. The DEX Metadata Consumer chooses the data element(s) mostly matching its needs and with the [QRPH-44] transaction (claiming the MPQ Document Type Binding Option) asks to the DEX Metadata Source about the Document Entry metadata describing the type of documents in the HIE system containing the data element(s) previously chosen.

After the Document Entry metadata have been retrieved, a [ITI-51] transaction to the MPQ Document Registry (in the HIE system) is sent by the MPQ Document Consumer (in the TTP) to retrieve the registry entries about the documents containing the “diagnosis” information. If also a specific period of time for the diagnosis was specified in the [PCC-1] Request, in the [ITI-51] transaction the $XDSDocumentEntryCreationTimeFrom and $XDSDocumentEntryCreationTimeTo query parameters are valued corresponding to respectively the beginning and end of the period of interest. A [ITI-39] transaction is sent then by the XCA Initiating Gateway in the TTP to retrieve the documents containing the “diagnosis” information from the HIE system.

The diagnosis information is then extracted by the TTP from documents together with other details about the diagnosis and the patient’s demographic information to be provided in the [PCC-1] Response. Since the researcher is allowed to have access only to anonymous data, in the [PCC-1] Response the patient’ identifier used in the HIE system has to be replaced by a wildcard (as “*”). Among all the “diagnosis events”, only the events related to “diagnosis of type 1 diabetes” have to be kept by the TTP and provided by the QED Clinical Data Source to the QED Clinical Data Consumer in the [PCC-1] Response.

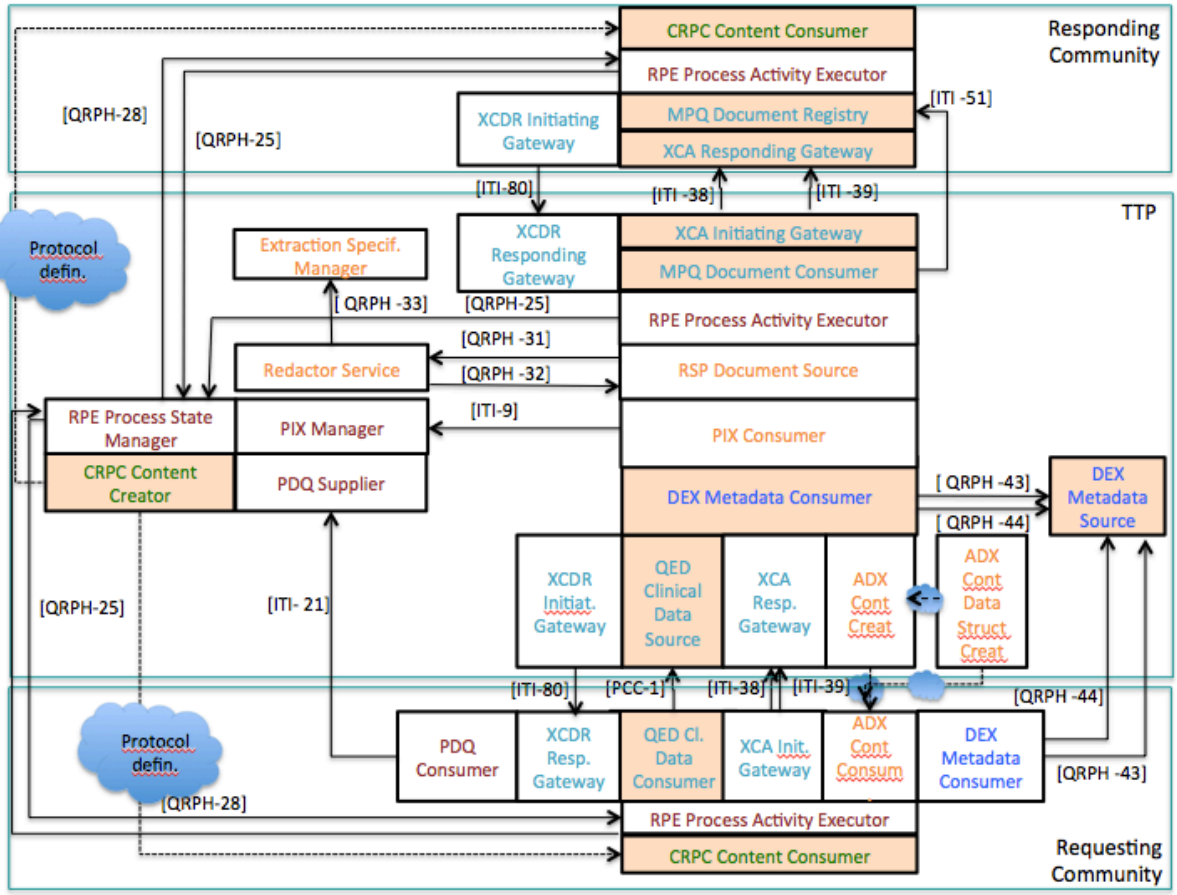


Figure 5.2-2: IHE profiles involved in use case 1

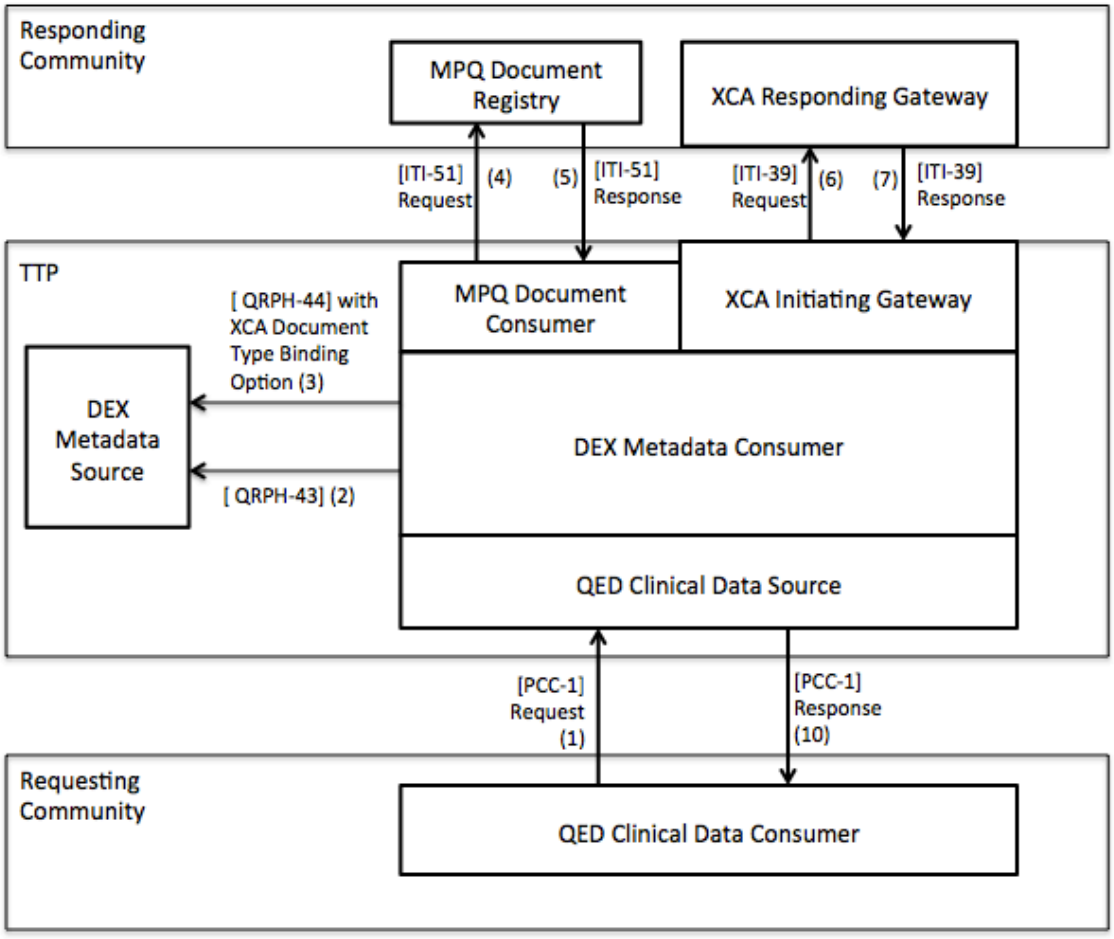


Figure 5.2-3: IHE profiles and workflow for specific for use case 1

5.3 Use case 2: CRFs retrieval for clinical purposes

The second use case is about clinicians working in the clinical community that would like to have access to data collected during clinical trials and stored in the research community. During clinical trials, lots of Case Report Forms (CRFs) have been administered to patients enrolled in the study, containing for example their clinical parameters values, main clinical events related to the study outcome, quality of life level, behavioral habits, psychological and social information.

An illustrative use case is the following one: a patient enrolled in a clinical trial (performed by the research organization) goes to his family doctor because of a thoracic pain. The patient tells the doctor about his participation in a clinical trial about a new drug meant to reduce anxiety. The doctor, who does not understand the cause of the pain, would like to have access to the patient's data collected during the trial, especially to CRFs with anxiety information not

available in the patient's EHR. However, in the current state, the doctor is not allowed to retrieve data stored in the research community.

The IHE standard technical solution that would allow the clinician to have access to CRFs available in the research community is represented in Figure 5.3-2 and Figure 5.3-3. Specifically, the Figure 5.3-2 highlights IHE actors to be implemented to solve this specific use-case, the Figure 5.3-3 shows only the specific actors involved in the process of query for data retrieval and the order of transactions flow.

First of all, the clinical community has to ask to the research organization the permission for clinicians to have access (in respect of legal and privacy issues) to documents collected during the clinical trial for patients they are treating. A protocol has to be defined and agreed by the parties (and approved by Data Protection Authority and/or Ethics Committee if needed) and created according to the CRPC Profile by the TTP where it is stored.

In this specific example, the protocol defines that the documents stored in the research organization can to be provided to the clinicians only as pseudonymized documents (the Trial ID used to identify the patient during the clinical trial has to be replaced with the subject ID used by the clinical community). Therefore, a linkage between the Trial ID and the subject ID has to be performed by the TTP in the initial phase of the study: each of the two communities have to provide to the TTP their patient's identifier and all the other demographic information related to each patient. A RPE [QRPH-25] transaction is sent by the RPE Process Activity Executor in both the two communities to provide (for each patient belonging to the specific community) to the TTP (playing the role of the RPE Process State Manager) the demographic information useful for the linkage. After and if a linkage is performed, the patient is "enrolled" in the study and a notification of the enrollment is sent by the RPE Process State Manager to both the two RPE Process Activity Executor actors with a RPE [QRPH-28] transaction.

After these preliminary steps have been concluded, the clinician can start to perform queries to the TTP in order to retrieve the CRFs (e.g., containing a diagnosis of anxiety) for the patient he is assisting. The conceptual schema of the query is indicated in Figure 5.3-1.

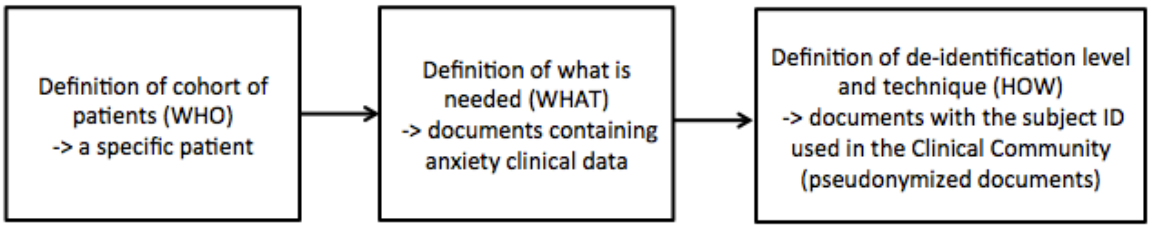


Figure 5.3-1: Query definition for use case 2

First of all, the clinician has to identify which is the "data element" available in the research organization that mostly matches the information he is looking for (in the specific example, a psychological disease diagnosis): to this purpose a [QRPH-43] transaction is sent by the DEX

Metadata Consumer played by clinician's EHR to the DEX Metadata Source played by TTP to retrieve the data element(s) that potentially match the "psychological disease diagnosis" information. The DEX Metadata Consumer chooses the data element(s) mostly matching his needs (for example the "PSYCO_PROBLEM" data element) and with the [QRPH-44] transaction (with a XCA Document Type Binding Option) asks to the DEX Metadata Source about the Document Entry metadata describing the type of document containing the data element(s) previously chosen (in this specific example the type of document containing the "PSYCO_PROBLEM" data element is the "Anxiety CRF").

After the Document Entry metadata retrieval, the actual query to retrieve the anxiety CRFs related to the specific patient (identified with the subject ID) can be initiated. The clinician's EHR forwards the request (e.g., using XDS queries or actors grouping) to the XCA Initiating Gateway in the clinical community to retrieve (from the XCA Initiating Gateway played by the TTP) the registry entries about the documents of interest. Before forwarding the request to the research organization, the PIX Consumer (played by the TTP) asks to the PIX Manager (played by another system in the TTP) about the Trial ID corresponding to the subject ID. This information is used to value the $XDSDocumentEntryPatientId query parameter in the [ITI-38] transaction Request sent by the XCA Initiating Gateway of the TTP to the XCA Initiating Gateway of the research organization: this query parameter specifies the patient ID for which the anxiety CRFs are looked for. The Response of the [ITI-38] transaction is then forwarded from the XCA Responding Gateway of the TTP to the XCA Initiating Gateway in the clinical community in the Response of the first [ITI-38] transaction.

After that, the XCA Initiating Gateway of the clinical community initiates a [ITI-39] transaction to the XCA Initiating Gateway of the TTP to retrieve the patient's anxiety CRFs. The request is then forwarded with another [ITI-39] transaction from the XCA Initiating Gateway of the TTP to the XCA Initiating Gateway of the research organization. After the TTP has retrieved the documents of interest (provided in the [ITI-39] Response), the process of document pseudonymization (to replace the Trial ID with the Subject ID) is then performed by the TTP. First, the PIX Consumer played the TTP asks to the PIX Manager played by another system in the TTP about the Subject ID corresponding to the Trial ID. After that, the TTP redacts the documents with RSP transactions ([QRPH-31], [QRPH-32], [QRPH-33]) and finally fills in the redacted document the blanked field containing the "patient ID" with the Subject ID.

The final pseudonymized document is finally provided in the [ITI-39] Response from the XCA Responding Gateway of the TTP to the XCA Initiating Gateway of the clinical community. Through XDS transactions (if the clinical community is organized as an XDS Environment), or actor grouping or other internal mechanisms, the subject ID's anxiety CRFs are finally provided to the clinician's EHR.

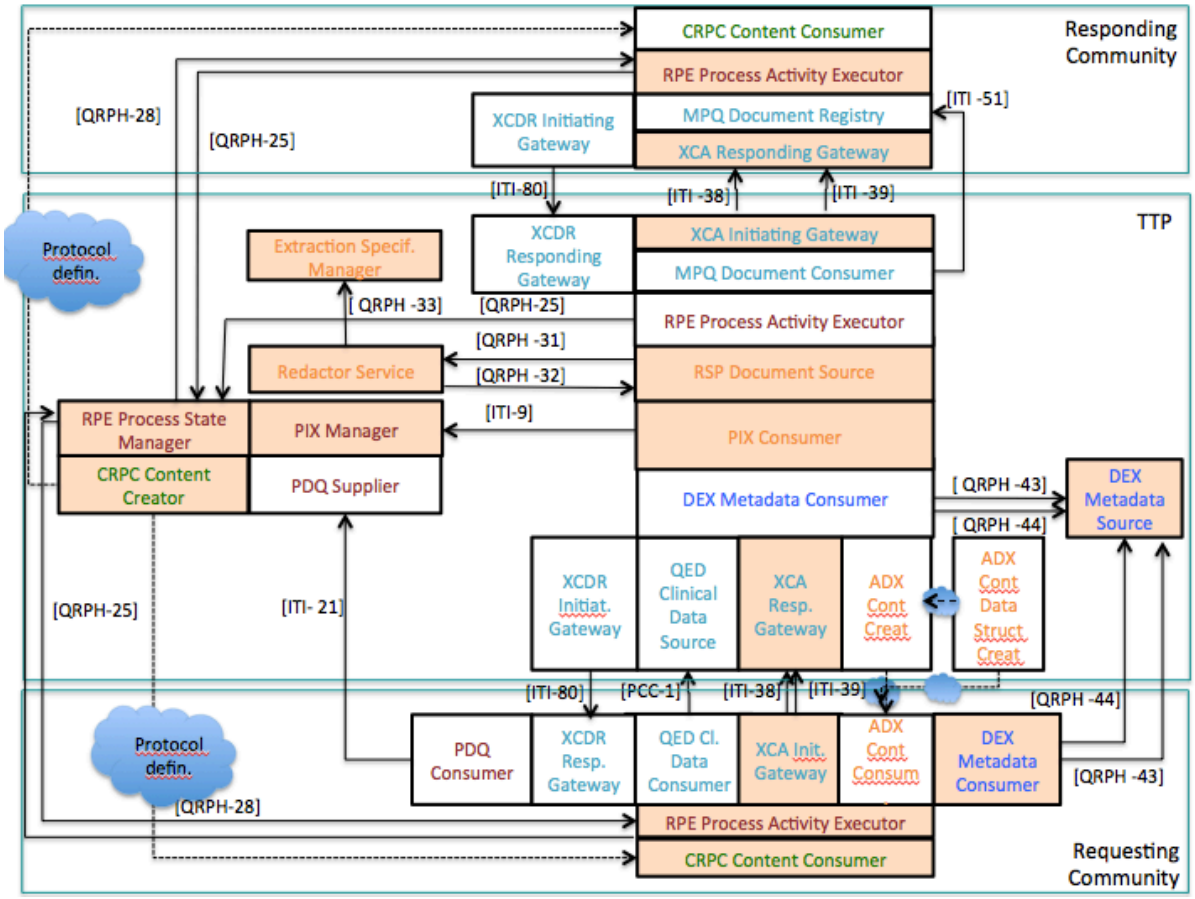


Figure 5.3-2: IHE profiles involved in use case 2

1810

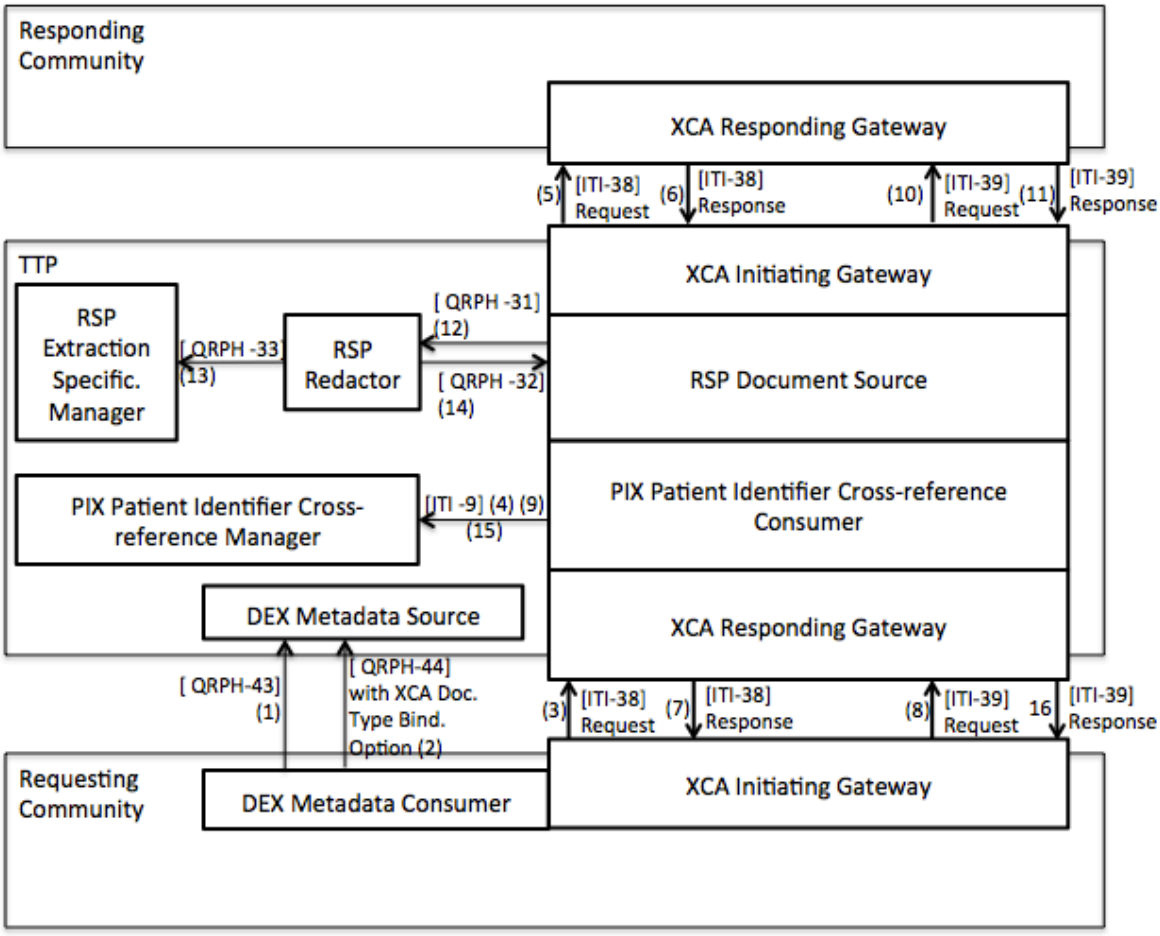


Figure 5.3-3: IHE profiles and workflow for specific for use case 2

1815

Appendix A – Privacy and ethics jurisdictional background

This informative appendix presents and discusses the main international standards about the secondary use of health data and focuses on the current EU legislation about this subject.

References:

- European Directive 95/46/EC
- European Directive 2013/37/UE
- Opinion 3/2013 of working Group Article 29
- Opinion 5/2014 of working Group Article 29
- Opinion 6/2013 of working Group Article 29
- ISO/TS 25237:2008 Health informatics -- Pseudonymization
- ISO/TS 14265:2011 Health Informatics - Classification of purposes for processing personal health information

Introduction

Health data constitutes a significant resource in most countries and it makes economic and ethical sense to use this data as much as possible to improve population health and the effectiveness and the efficiency of health care systems. Central to the assessment of both the health of populations and the quality and efficiency of health care services are data to measure, monitor and compare performance.

Regional, national and international reports on health and health care are entirely dependent upon monitoring policies and investments in data infrastructure that either facilitate or restrict data and analysis.

Understanding the progress of the health of populations and understanding the performance and quality of health care systems requires the ability to monitor the same individuals over time, as they experience health care events, receive treatments, experience improvements or deteriorations in their health and live or die. It also requires understanding the distribution of health and health outcomes across different groups in the population and understanding variations in care quality and health outcomes.[7]

Re-use of data

Secondary use of data occurs when data is used for a purpose different from the purpose for which the data was initially collected. Enabling secondary use of medical data by healthcare professionals and researchers is important to improve the quality of health care and research effectiveness. At the same time, it is important to protect patient privacy and to ensure that no harm is done to a patient through the use of the data.

7 secondary analysis of health data to generate health care quality information_ http://www.garanteprivacy.it/documents/10160/2052659/1895987

The debate on the re-use of information held by the public sector which is mainly directed to the re-use of public data, and not of personal data. However, already in the Green Paper on public sector information in the Information Society of the European Commission it has taken into account the need for protection of privacy in the case in which the archives and public records contain personal data (pag.108-112). In fact, a part of public information has a personal nature, think of the population, company, vehicle or credit, employment or social welfare.

On the same topic the Article 29 Working Party (WP29) issued its opinion n. 6 of June 5, 2013 named "Opinion 6/2013 on open data and on the reuse of public sector information ("PSI")". This Opinion followed the adoption of Directive 2013/37/UE[8] of European Parliament and Council amending Directive 2003/98/EC[9] on the re-use of public sector (the “PSI Directive”). The aim of Opinion 6/2013 is to help ensuring a common understanding on the applicable legal framework, and to offer consistent guidance and best practice examples on how to implement the PSI Directive (as amended) with regard to the processing of personal data.

WP29 underlines that the lack of a consistent approach may weaken the position of the individuals concerned. It may also impose unnecessary regulatory burdens for businesses and other organizations operating cross-borders and thus represent an obstacle to develop a common European market for re-use. On one hand, data subjects must be assured that their data will be consistently protected irrespective of their transfer to another Member State, for the purposes of re-use. On the other hand, undue complexity and fragmentation should be avoided also to enable the free flow of personal data across Europe, which represents another key objective of Directive 95/46/EC[10].

The WP29 emphasizes the necessity of adhering to the principles of “data protection by design and by default” and to ensure that data protection concerns are addressed at an early stage. In particular, the WP29 strongly recommends to public bodies to carry out a data protection impact assessment before making available personal data for reuse. Member States should also consider making such an impact assessment mandatory under national legislation or promoting it as a best practice. In any case, this should happen prior to the disclosure of information and to the decision of making it available for re-use, even if it is not expressly envisaged by national laws.

The assessment should also establish a legal basis for sensitive data disclosure (and potential legal basis for reuse); moreover, it should identify the principles of purpose limitation, proportionality and data minimization, and consider the special protection required for sensitive data. In carrying out this evaluation the potential impact on the data subjects should be carefully considered.

Data protection laws do not usually allow that public bodies publicly disclose personal data collected for another purpose. Thus, in these cases their reuse as part of PSI reuse initiatives is not possible. Rather than personal data, it is typically statistical data derived from personal data

8 http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF

9 http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:en:PDF

10 http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML

that is and that should, in principle, be made available for reuse. This is the most effective solution to minimize the risks of inadvertent disclosure of personal data. These anonymized and aggregated datasets should not allow re-identification of individuals and therefore should not contain personal data.

The Green Paper mentioned above identifies the Directive on Data Protection Directive 95 /46 / EC as the necessary reference point for the protection of privacy both for public and private sector, and states that the "competent public bodies are in charge of conciliating on one hand the need for open access (for commercial purposes or others) and on the other hand the right to protection of privacy, by applying the principles established in the EC Directive, in particular that of purpose limitation."

The European Commission (EC) realized that this diversity of national legislation impedes uniform data protection and the free flow of data within the EU zone. Therefore the EC drafted Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data ('Directive') to harmonize data protection regulation within the EU. The Directive regulates the processing of personal data and the free movement of such data and had to be implemented into national law by the end of 1998. Currently all member states have implemented it into their own national data protection legislation. The Directive is not a 'closed regulatory system'; it leaves open a certain scope for policy making at national level, however certain minimum requirements must be met.

Starting from the assumption that it is illegal and therefore forbidden to process personal data that can identify the person to which they refer in a manner inconsistent with the purpose stated at the time of their collection, there is the need to pay particular attention to the purposes for which the data is reused. On this topic WP29 as well as the European Directive 95/46 / EC expressed their opinion.

Purpose of the secondary data usage

Article 6 sub b) European Directive 95/46/EC states that personal data can only be processed for specified explicit and legitimate purposes and may not be processed further in a way incompatible with those purposes. Further processing of data for historical, statistical or scientific purposes shall not be considered as incompatible provided that member states provide appropriate safeguards.

Personal data may be processed only under the following circumstances (art. 7):

- when the data subject has given his consent; or
- when the processing is necessary for the performance of a contract to which the data subject is party or to take steps at the request of the data subject prior to entering into a contract; or
- when processing is necessary for compliance with a legal obligation to which the controller is subject; or
- when processing is necessary to protect the vital interests of the data subject; or

- when interest or in the exercise of the official authority vested in the controller or in a third party, processing is necessary to perform a task carried out in the public to whom the data are disclosed; or
- processing is necessary for the legitimate interests pursued by the controller or by the third party(s) to whom the data are disclosed, except when overridden by the interests for fundamental rights and freedoms of the data subject.

Personal data must be processed fairly and lawfully.

Personal data must be collected for specific and legitimate purposes and not further processed in a way incompatible with those purposes. Further processing of data for historical, statistical or scientific purposes shall not be considered as incompatible provided that member states provide appropriate safeguards.

Personal data must be adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed.

Personal data must be kept in a form, which permits identification of data subjects for, no longer than is necessary for the purposes for which the data were collected or for which they are further processed. Member states shall lay down appropriate safeguards for personal data stored for longer periods for historical, statistical or scientific use.

It is important to note that any purpose must be specified, that is, sufficiently defined to enable the implementation of any necessary data protection safeguards, and to delimit the scope of the processing operation.

Further, to be explicit, the purpose must be sufficiently unambiguous and clearly expressed. Comparing the notion of ‘explicit purpose’ with the notion of ‘hidden purpose’ may help to understand the scope of this requirement.

Finally, purposes must also be legitimate. This notion goes beyond the requirement to have a legal ground for the processing under Article 7 of the Directive and also extends to other areas of law.

An important aspect to consider is the assessment of compatibility, as established by the WP29 in his opinion 3/2013: personal data collected for one or more purposes shall 'not be further processed in a way incompatible with those purposes.

Key factors to be considered during the compatibility assessment:

- the relationship between the purposes for which the data have been collected and the purposes of further processing;
- the context in which the data have been collected and the reasonable expectations of the data subjects as to their further use;

- the nature of the data and the impact of the further processing on the data subjects (the more sensitive the information involved, the narrower the scope for compatible use would be);
- the safeguards applied by the controller to ensure fair processing and to prevent any undue impact on the data subjects. Appropriate additional measures could thus, in principle, serve as 'compensation' for a change of purpose or for the fact that the purposes have not been specified as clearly in the beginning as they should have been. This might require technical and/or organizational measures to ensure functional separation (such as partial or full anonymization, pseudonymization, and aggregation of data), but also additional steps taken for the benefit of the data subjects, such as increased transparency, with the possibility to object or provide specific consent. Whether the result is acceptable will depend on the compatibility assessment as a whole (i.e., including those measures and their effect on the other aspects mentioned above).

Opinion 3/2013 adopted by WP29 analyses the principles linked to purpose limitation to protect data subjects by setting limits on how data controllers are able to use their data and at the same time by offering some degree of flexibility for data controllers. The concept of purpose limitation has two main building blocks: personal data must be collected for 'specified, explicit and legitimate' purposes (purpose specification) and not be 'further processed in a way incompatible' with those purposes (compatible use).

WP29 establishes that further processing for a different purpose does not necessarily mean that it is incompatible: compatibility needs to be assessed on a case-by-case basis. A substantive compatibility assessment requires an assessment of all relevant circumstances.

Processing of personal data in a way incompatible with the purposes specified at collection is against the law and therefore prohibited.

In order to strengthen the protection of Personal Information Health treated, stored and transmitted by ICT tools and afterwards used by doctors and other health specialists, in 2011 the' International Organization for Standardisation (ISO) published the Technical Specification ISO / TS 14265 "Health informatics – Classification of purposes for processing personal health information". The document identifies a classification system that indicates when such information may be treated.

The ISO / TS 14265 provides a framework to classify how to use the information on the basis of specific needs and different actors (health organizations, regional health authorities, health services); the goal is to facilitate the systematic management of information in health services and communication of EHRs across organizational boundaries and jurisdiction.

Data anonymization

The term "anonymization" refers to data that can no longer be considered personal upon Article 2, letter a) of Directive 95/46 / EC. Upon the above Article 2, letter a), the term "personal data" is intended as "any information relating to an identified or identifiable individual ("data subject") while an identifiable person is someone who can be identified, directly or indirectly, in particular

through an identification number or one or more factors referring to specific physical, physiological, mental, economic, cultural or social identity". Recital 26 of Directive 95/46/EC is also relevant for this purpose and states that "to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any others to identify the said person".

Directive 95/46/EC refers to anonymization in Recital 26 to exclude anonymized data from the scope of data protection legislation: “Whereas the principles of protection must apply to any information concerning an identified or identifiable person; whereas, to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person; whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable; whereas codes of conduct within the meaning of Article 27 may be a useful instrument for providing guidance as to the ways in which data may be rendered anonymous and retained in a form in which identification of the data subject is no longer possible”.

Regarding anonymization, the starting assumption is that the personal data must have been collected and processed in compliance with the applicable legislation on the retention of data in an identifiable format. In this context, the anonymization process, intended as the processing of such personal data to achieve their anonymization, represents a “further processing”. As such, this processing must comply with the test of compatibility in accordance with the guidelines provided by the Working Party in its Opinion 03/201311 on purpose limitation.

This means that, in principle, the legal basis for anonymization can be found in any of the grounds mentioned in Article 7 (including the data controller’s legitimate interest) of European directive 95/46/EC provided the data quality requirements of Article 6 of the Directive mentioned above.

The anonymized data should be distinguished from data that has been manipulated using various techniques to reduce the risk of re-identification of the persons concerned, but without reaching the threshold set by Article 2, letter a) and recital 26 of Directive 95/46/EC. In many situations, these techniques are appropriate only if the diffusion is limited to the purpose of re-use by third parties subject to control, but not in the case of public dissemination and with open license reuse.

In relation with anonymization it is useful to analyze Opinion5/2014 of WP29 on anonymization techniques. In this Opinion the effectiveness and limits of existing anonymization techniques is analyzed taking into consideration the legal background of data protection. In this document it is established that anonymous data must be distinguished from the data that have been manipulated using various techniques to reduce the risk of re-identification of the persons concerned, but without reaching the threshold set by Article 2, letter a) and recital 26 of Directive 95/46/EC.

[11] parere del Gruppo di lavoro articolo 29 http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

In many situations, these techniques are appropriate only if the diffusion is limited to the purpose of re-use by third parties subject to control, but not in the case of public dissemination and with open license reuse.

In order to evaluate the robustness of each anonymization technique, Opinion 5/2014 identifies the following three criteria:

- is it still possible to identify an individual?
- is it still possible to link records related to an individual?
- can information on an individual be inferred?

Being aware of the main strengths and weaknesses of each technique helps to choose how to design an adequate anonymization process in a given context.

In this context WP29 considers pseudonymization and clarifies how it is not an anonymization method. Pseudonymization is not a method of anonymization. It merely reduces the linkage of a dataset with the original identity of a data subject, and is accordingly a useful security measure. Pseudonymization consists of replacing one attribute (typically a unique attribute) in a record by another. The natural person is therefore still likely to be identified indirectly; accordingly, pseudonymization will not result in an anonymous dataset if used alone. Nevertheless, the issue is discussed in this Opinion because of the many misconceptions and mistakes on its use.

The most used pseudonymization techniques are encryption with secret key, hash function, salted-hash function, keyed hash function with stored key, deterministic encryption or keyed-hash function with deletion of the key.

With regard to pseudoniymisation we need to talk about ISO / TS 25237: 2008.

In theme of pseudoniymisation ISO / TS 25237: 2008. ISO/TS 25237:2008 contains principles and requirements for privacy protection using pseudonymization services for the protection of personal health information. ISO/TS 25237:2008 is applicable to organizations that make a claim of trustworthiness for operations engaged in pseudonymization services.

ISO/TS 25237:2008 is applicable to organizations that make a claim of trustworthiness for operations engaged in pseudonymization services, which may be national or trans-border. It will serve as a general guide for implementers, as well as for quality assurance purposes, assisting users to determine their trust in the services provided. Application areas include, but are not limited to:

- Research, or other secondary use of clinical data
- Clinical trials and post-marketing surveillance
- Public health monitoring and assessment
- Confidential patient-safety reporting (e.g., adverse drug effects)
- Comparative quality indicator reporting

- Peer review
- Consumer groups.

ISO/TS 25237:2008 was developed by ISO technical committee ISO/TC 215, Health informatics. It provides a conceptual model of the problem areas, requirements for trustworthy practices, and specifications to support the planning and implementation of pseudonymization services.
Regarding the data anonymization, WP29 again with the opinion 6/2013, poses the question on who should carry out the aggregation and data anonymization and when it should be done.

According to WP29 both aggregation and anonymization should occur at the earliest opportunity – by the data controller or by a trusted third party acting on behalf of the controller or several controllers (and who is also in possession of the necessary specialized skills). It cannot be left to the re-user to carry out the anonymization, for example as a licensing condition. Further, it is important to ensure that the possible third party organization carrying out the aggregation and anonymization has no conflict of interest and is clearly held accountable that the personal data will only be used to carry out the anonymization and that all the necessary safeguards are put in place to this effect. The third party should also be able to guarantee that the personal data from which the aggregated and anonymized datasets are derived should be deleted as soon as they are no longer required for that purpose.

Another important aspect that should be considered relates to the inability to make anonymous the data pursuant to Article 2 of Directive 95/46/EC. In this case, you must continue to use the data in accordance with the provisions of the regulations regarding data protection.

What above described is of course based on the existing legislation. However, on 14th April 2016 the European Parliament approved the new EU General Data Protection Regulation that must be applied in Member States within 2 years. From an operational point of view, in next 2 years it will be possible to analyze if this new regulation substantially modify the practices established by WP29 and currently applied.

Appendix B – IHE profiles for further privacy and security issues

IHE published the Access Control White Paper (http://ihe.net/Technical_Framework/upload/IHE_ITI_TF_WhitePaper_AccessControl_2009-09-28.pdf) that defines guidelines to build an Access Control system. That document identifies main Security issues related to Clinical Data exchange. With respect to the prevention of inappropriate or illegal disclosure, it is crucial that providers of medical data can be sure that data consuming parties enforce access constraints conformant to the purposes under which that data was provided. Therefore the definition and enforcement of access rules for medical data and services throughout workflows is a precondition for any cooperative patient treatment. This is especially true in the case of secondary data usage.

