Security Working Group

Bulk Data Transfer Privacy Requirements

Mohammad Jafari,
Kathleen Connor, John M. Davis, Christopher Shawn

Version 1.5

June 28, 2019

(revised for publication on 11/13/2019)

Table of Contents

List of Figures

List of Tables

1 INTRODUCTION

This report presents and discusses the privacy requirements for the emerging Bulk Data Transfer (BDT) specifications for the Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR). The requirements formulated in this report have emerged as a result of the following:

- Privacy analysis of the draft specifications for the FHIR BDT;
- Engaging with the developer community via the FHIR community forum;
- Demonstration and presentation of Privacy-Aware Bulk Data Transfer at the HL7 FHIR Connectathon in May 2019 in Montreal;
- Discussions with other stakeholders such as developers and members of HL7 Security and Security and Community-Based Care and Privacy (CBCP) Working Groups at the HL7 FHIR Connectathon and subsequent Work Group meetings in May 2019 in Montreal.

Note that the FHIR BDT specification and its implementations are still in their early phases and as more systems implement the specifications and new applications and use-cases emerge, these requirements may need to change or be updated.

1.1 Scope of This Report

This report focuses on requirements broadly applicable to bulk data transfer across different applications and use-cases, including security, privacy, data integrity, and provenance requirements. Note that each application area in which bulk data transfer is used, often has further security and privacy requirements which are specific to that application area and will not be covered in this report.

1.2 Bulk Data Transfer

The goal of BDT is to provide an efficient mechanism for transferring health information in bulk where making individual FHIR API calls is inefficient because of the synchronous nature of the FHIR API calls, potential large volume of the data, and other logistic reasons such as rate-limiting or transaction overheads for API calls.

Moreover, these specifications are intended to provide a standard FHIR-based alternative to existing proprietary mechanisms for bulk import/export into/from electronic health record (EHR) systems.

Existing and potential use-cases of BDT include exporting and importing data in applications such as: analytics, machine learning and data mining, clinical data warehouse for research and cohort identification, population-level health reports and notifiable disease submissions, billing and claims processing, collecting data from offline medical devices, or cross-EHR exchanges.

1.3 FHIR Bulk Data Transfer Draft

The BDT is a draft specification [1] for an Application Programming Interface (API) which enables reading or writing a large number of resources from/to a FHIR server. There are two parts to the BDT API:

- The *import* API enables a client to send a potentially large volume of health information to a FHIR server in a single transaction. Imported data can originate from another FHIR server, a proprietary electronic health record (EHR) system, or other sources such as a medical device.

- Conversely, the *export* API enables clients to specify a broad set of FHIR resources and download them *en masse* as a single (or a few) file(s). Exported data can be sent to another FHIR server, a proprietary EHR system, a clinical data warehouse system, or other types of storage and processing systems.

2 REQUIREMENTS

The privacy requirement for BDT are classified into the following four groups. Figure 1 and Figure 2 depict the interplay of these four types of requirements in bulk export and import operations:

1. *Authorization* requirements specify control over whether or not a client's request for bulk import or export should be permitted.
2. *Filtering* requirements specify, at a more fine-grained level, what resources will appear in the results of an export or accepted in an import operation based on the authorization level of the client. These are different from authorization requirements in that, rather than rejecting the client's request, the response is modified at a fine-grained level to match the client's authorization level.
3. *Transformation* requirements specify the requirements for applying functions on imported or exported resources to modify and transform the content of the resource, to comply with privacy and security requirements.
4. *Provenance* requirements specify the recording and consumption of provenance information in an export or import operation.

The requirements in each of these categories will be discussed in the rest of this section. Table 1 at the end of this report, provides a single-page summary of all the requirements.

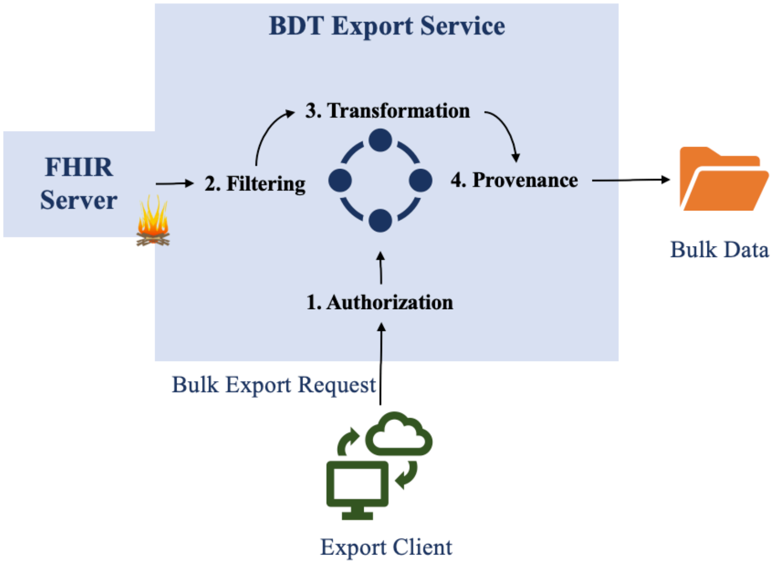

Figure 1: The interplay of the four groups of requirements in bulk export of FHIR resources

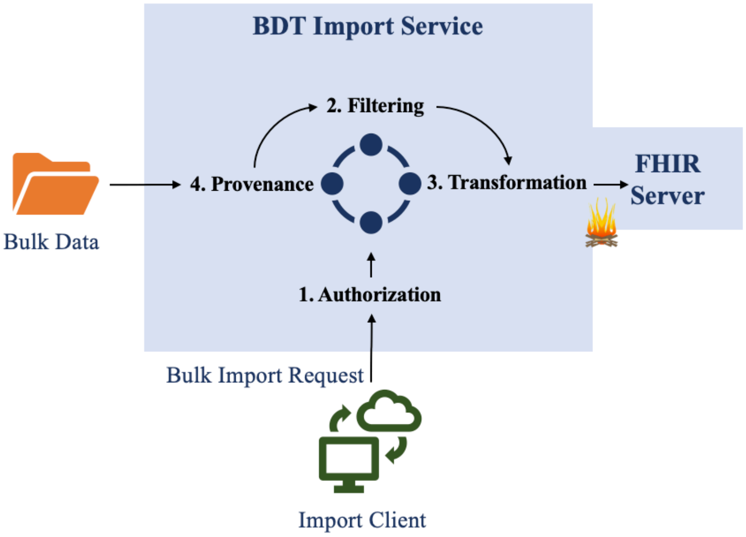

Figure 2: The interplay of the four groups of requirements in bulk import of FHIR resources

2.1 Authorization

Authorization requirements specify how the BDT should exercise control over whether or not a client's request for bulk import or export is permitted. This section only focuses on authorization requirements specific to BDT; other general authorization requirements, such as enforcing policies based on client identifier, purpose of use, or other attributes are not re-iterated here.

2.1.1 Explicit Bulk Permissions

The BDT server SHALL require, based on policy, that a client possess an explicit permission (clearance) for bulk access before permitting a bulk request by that client.

This requirement emphasizes the distinction between regular read/write permissions granted to clients and the broader bulk export/import permissions. While the export/import permission can, in theory, be reduced to (a potentially large) number of read/write operations, it is important to recognize that bulk requests are a different type of transaction as they give the client a different level of power to read or write data at a larger scale with a relatively small audit footprint.

The analytical capabilities resulting from a client's bulk access to health information (e.g., correlation of patient identifiers, profiling, and inference via data mining and machine learning techniques) make this type of access distinct from individual transactions focused on limited information belonging to a single patient. Some jurisdictions have already recognized these types of processing to be distinct from individual access; for example, the European Union's General Data Protection Regulation (GDPR) requires data subjects' consent before being subject to decisions resulting from automated data processing and profiling [2] and additional measures including Data Protection Impact Assessment (DPIA) for systematic and extensive or large-scale processing of sensitive or personal information [3]. The Dutch Data Protection Authority has specifically designated *large-scale electronic exchange of health data* as well as any large-scale processing of genetic and health information as a *high-risk* category which requires Data Protection Impact Assessment (DPIA); this includes large-scale processing of health information by healthcare facilities, insurers, social and reintegration services, and research institutions [4].

2.1.2 Resource-Type-Specific Permissions

The BDT service SHOULD require, based on policy, that the client possess specific bulk permissions (clearances) granting access for the requested resource types.

This requirement ensures that the BDT service is capable of enforcing granular access control on bulk access requests based on the *types* of requested FHIR resources. For example, a client which is only granted permission for bulk export of Immunization resources must not be permitted to export *all* patient resources, or any other resource types not explicitly authorized, such as Medication Requests.

For efficiency and usability, the BDT service shall recognize and support mechanisms for granting permissions for bulk access to a set of, or if needed, *all* resource types. This ensures that a client which needs to request bulk export/import bundles of multiple types are resources (often related resources) can do that without having to re-attempt authorization or send a separate request for each resource type.

2.2 Filtering

Filtering requirements address the BDT service's fine-grained control over what resources will appear in the results of an export request, or accepted from the offered resources in an import request.

These requirements focus on the ability of the BDT service to enforce more specific parts of its privacy policies (e.g., such as compatibility of resource security labels with the client's clearances) without forcing the client to explicitly refine the request –which could be privacy-revealing by exposing, via inference, the existence of sensitive information or some details of the privacy policies (e.g., patient consents).

For example, when a client, which has previously had access to exporting Immunization records, gets a rejection from the BDT service in response to an export request, it can infer that this may be because some *restricted* Immunizations have emerged in the result of the export query, especially if a more refined query which narrows down the requested information to non-restricted resources, is accepted. Accepting the query and leaving out the *restricted* resources from the results in this case enables the BDT service to fulfill the request without divulging the existence of these *restricted* resources to the client.

2.2.1 Filtering Based on Security Labels in Export

Based on policies, the BDT service SHOULD filter the results of an export request, based on security labels, by matching resource labels with the client's clearance.

For example, if, based on policy, a client is granted only the permission to export Immunization resources with *normal* confidentiality, the BDT service must exclude any FHIR resource marked with higher levels of confidentiality (e.g., *restricted*) from the export results. The record of such filtering may be communicated to the client if policies permit.

2.2.2 Filtering Based on Security Labels in Import

Based on policies, the BDT service MAY filter the submitted resources in an import request, based on security labels, by matching resource labels with the client's or the service's clearance. For example, the BDT service for a marketing organization can skip any resources specifically marked with purposes other than marketing.

2.2.3 Filtering Based on Patient Consent

When overarching policies give the patient the power to opt-out or require the patient's explicit opt-in, the BDT service SHALL filter the results of an export request based on the patient's consent.

For example, if overarching policies allow the patient to opt out of sharing their health information for the purpose of research, the BDT service must be able to exclude the information belonging to any patient who has opted out (based on the consent) when a client requests a bulk export for the purpose of *research*. The record of applied filters may be communicated to the client if policies permit that.

2.3 Transformation

Transformation requirements focus on applying functions on the imported or exported resource which modify and transform their content.

2.3.1 Security Labeling in Export

Based on policies, the BDT service SHOULD add security labels to exported resources or MAY update existing security labels.

The BDT service must be able to invoke labeling or re-labeling of security labels on the outgoing resources in an export, depending on the policies, client attributes, and the transaction context. This is often by leveraging an existing Security Labeling Server (SLS). Re-labeling for a specific client may be necessary, for example, when particular handling instructions (e.g., no-redisclosure) must be added based on the identity of the client and the stated purpose of use for the transaction.

2.3.2 Security Labeling in Import

Based on policies, the BDT service MAY add security labels to imported resources or update existing security labels.

The BDT service must be able to arrange adding security labels or changing/updating existing labels on the offered resources for an import based on policies, depending on the policies, client attributes, and the transaction context. This is often by leveraging an existing Security Labeling Server (SLS). This may be necessary when the labels do not exist to begin with, or there is a need to relabel the incoming data based on policy stipulations.

2.3.3 Integrity Labeling for Import

Based on policies, the BDT service SHOULD arrange assigning integrity labels to imported data depending on its level of trust in the client and the origin of the data.

Note that although integrity labels are a type of security label, this specific requirement is singled out to highlight its importance. The level of confidence in an external source of data may have to be reflected with integrity labels when data is imported into a FHIR server. For example, data imported from a medical device in possession of the patient may need to be labelled with the *device-asserted* integrity label to reflect the degree of certainty and trust in the data. Assigning these labels often takes place by leveraging the SLS and may take into account the provenance information provided by the source of information.

2.3.4 Profile-Based Content Transformation in Export

Based on policies, the BDT service SHOULD apply transformation functions (detailed in standard profiles) on the content of exported resources.

This enables the BDT service to exercise fine-grained control over the content of the exported resources. Some examples of such transformation are:

- Anonymization and pseudonymization profiles that define functions for removing demographic and personally-identifiable attributes from resources.
- Summarization profiles that define functions to replace the level of specificity of an attribute, for example, by replacing a full name with initials, or by replacing the exact date of birth with the year or decade of birth.
- Masking profiles that define functions for replacing the value of certain attributes with an encrypted version only accessible to the entities with the decryption key.

A standard vocabulary for defining and referencing transformation functions which can be referenced in policies would enable a policy-based mechanism for applying such transformation functions. For example, a policy may require that the data exported for the purpose of *research* by client *x* be transformed by applying particular anonymization and summarization functions.

2.3.5 Profile-Based Content Transformation in Import

Based on policies, the BDT service MAY apply transformation functions (detailed in standard profiles) on the content of imported resources.

For example, the BDT service may apply an unmasking function on the content of the data offered for import, if resource content or specific attributes have been previously masked by ann equivalent function.

2.4 Provenance

Provenance requirements specify the capabilities for recording and consuming provenance information in an export or import operation.

2.4.1 Generating Origin Provenance for Export

The BDT SHOULD record the circumstances of the export in the form of a provenance resource accompanying the result of the export operation.

The circumstances of export may include information about the originating FHIR server or servers (if the data is collected from multiple back-end FHIR services), the transaction context, signatures, and other relevant information, such as applied filtering or content transformation functions. The details of how extensive this provenance record needs to be and what it should record are defined by applicable policies.

The provenance information will be consumed by the client or other entities on the client's side; it enables them to develop confidence in the origin of the exported data and the context in which they have been released.

2.4.2 Consuming Origin Provenance in Import

The BDT SHOULD recognize and consume the provenance information provided by the origin of the data at the time of import.

The provenance information generated by the origin BDT service at the time of export includes important information about the origin of the data and the context of its release; so, if the data is imported to another FHIR Server, the receiving BDT must be able to consume and incorporate this information. This can lead to other policy-based decisions by the BDT service, such as application of integrity labels or transformation functions (e.g., unmasking).

2.4.3 Generating Origin Provenance for Import

The BDT SHOULD record the circumstances of the import in the form of a provenance resource.

This is the BDT's own account of the circumstances of the import, which could incorporate the information from the original provenance provided by the client but may also include other information from the BDT's own perspective. For example, the identity of the client submitting the import or the BDT's level of trust in that client are information which may be either unavailable in the original provenance or different from what was recorded by the origin.

2.4.4 Transformation and Filtering Provenance

In the event of any transformation or filtering, the BDT MAY record provenance information regarding the transformations applied to the content of resources in import/export.

This information can be either incorporated into the generated record of import/export or be recorded in the form of a separate provenance resource. For example, if the BDT service applies a particular anonymization function based on a standard method, this can be either recorded in the provenance resource which captures the context of export, or as a separate provenance resource which captures the transformations applied to the exported resources, linked to the general export provenance.

Note that informing the client about the application of filtering on exported resources may be breach privacy since it may reveal to the client the existence of certain information that the client was not authorized to see. Thus, a careful analysis of the privacy consequences and ultimately the applicable policies will determines whether the BDT should record instances of resource filtering.

3 SYSTEM VIEW OF REQUIREMENTS

Figures 3 and 4 provide a system view of the requirements in the context of interactions with other privacy-related services, Access Control Service (ACS), Security Labeling Service (SLS), Privacy-Preserving Services (PPS), Provenance Service, and Consent Service.

This shows that the requirements in this report are at the cross-section of the BDT service with other respective privacy-related services and captures the interactions with these services. Ultimately, at the implementation-level, the BDT service invokes these existing privacy-related services to implement the requirements as discussed in this report, so, these requirements affect these privacy-related services as well.

3.1 Import Walk-Through

Upon receiving an import request, the BDT service typically performs the following steps (based on the numbers in Table 1, the corresponding requirement is referenced in parentheses):

- The BDT service invokes the ACS to authorize the import client and ensure the client is authorized to make bulk requests (1.1) and the imported resource types match the resource types which the client has been authorized to import (1.2).
- If the request is authorized, an import query processor inspects the client's query and the submitted data.
- The BDT service invokes the Provenance Service to inspect and consume the provenance information provided by the client (4.2) and generate its own provenance record for the import (4.3).
- Per policy and if applicable, the BDT service may invoke the SLS and PPS to filter the submitted resources and, based on the security labels, only accept resources for which the client has sufficient clearance and for which the server finds itself authorized (2.2). This event may be recorded in the form, or as part of, a provenance resource (4.4).
- Per policy and if applicable, the BDT invokes the SLS to label or re-label the imported data (3.2), particularly assigning integrity labels reflecting the origin and confidence in the imported data (3.3). This event may be recorded in the form, or as part of, a provenance resource (4.4).
- If applicable and based on policy, the BDT service may also invoke the PPS to apply transformations on the content of the resources (3.5). This event may be recorded in the form, or as part of, a provenance resource (4.4).

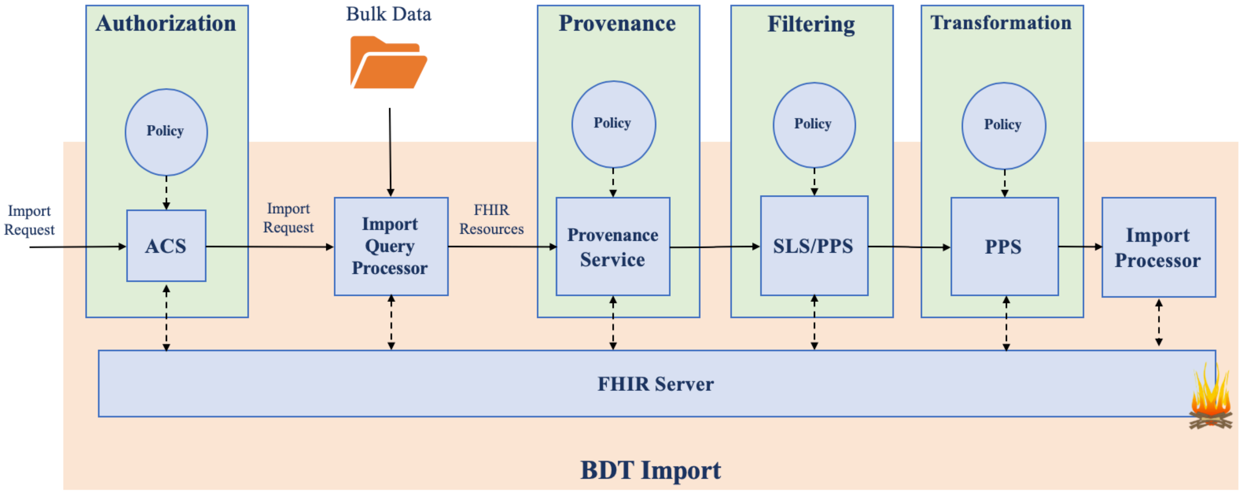

Figure 3: System view of the import requirements in the context of interactions with other privacy-related services

3.2 Export Walk-Through

Upon receiving an export request, the BDT service typically performs the following steps (the corresponding requirement is cited in parentheses based on the numbers assigned in Table 1):

- The BDT service invokes the ACS to authorize the export client and ensure the client is authorized to make bulk requests (1.1) and the export resource types match the resource types which the client has been authorized to export (1.2).
- If the request is authorized, an export query processor inspects the client's query and identifies the data to be exported.
- Per policy, the BDT service invokes the SLS and PPS to filter the candidate resources for the export and, based on the security labels, only allows resources for which the client has sufficient clearance (2.1). Moreover, if applicable, the BDT service uses the Consent Service to identify and filter out the resources belonging to patients who do not have sufficient consent for the transaction in question (2.3).
- Per policy and if required, the BDT service invokes the SLS to label or re-label the exported data (3.1).
- If applicable and based on policy, the BDT service may also invoke the PPS to apply transformations on the content of the resources (3.4). This event may be recorded in the form of a provenance resource (4.4).
- The BDT service invokes the Provenance Service to generate a provenance record for the export (4.1).

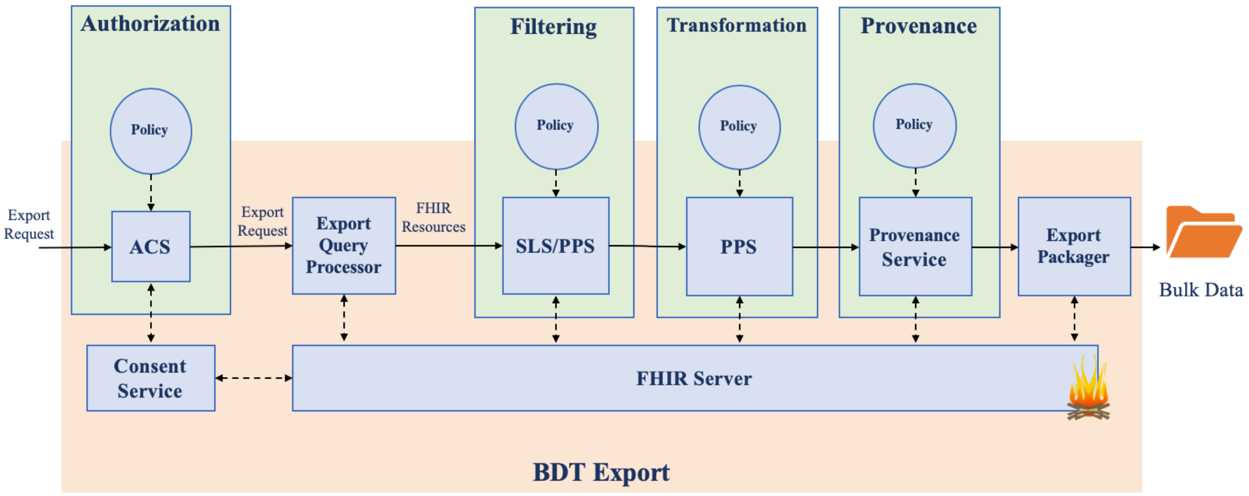

Figure 4: System view of the export requirements in the context of interactions with other privacy-related services

4 SUMMARY

Table 1: Summary of Requirements for FHIR Bulk Data Transfer

Title	Description	Scope	Services
1. Authorization			
1.1. Explicit Bulk Permissions	The BDT server SHALL require, based on policy, that a client possess an explicit permission (clearance) for bulk access before permitting a bulk request by that client.	Import, Export	ACS
1.2. Resource-Type-Specific Permissions	The BDT service SHOULD require, based on policy, that the client possess specific bulk permissions (clearances) granting access for the requested resource types.	Import, Export	ACS
2. Filtering			
2.1. Filtering Based on Security Labels on Export	Based on policies, the BDT service SHOULD filter the results of an export request, based on security labels, by matching resource labels with the client's clearance.	Export	SLS, PPS
2.2. Filtering Based on Security Labels on Import	Based on policies, the BDT service MAY filter the submitted resources in an import request, based on security labels, by matching resource labels with the client's clearance.	Import	SLS, PPS
2.3. Filtering Based on Patient Consent	When overarching policies give the patient the power to opt-out or require the patient's explicit opt-in, the BDT service SHALL filter the results of an export request based on the patient's consent.	Export	SLS, PPS, Consent Service
3. Transformation			
3.1. Security Labeling on Export	Based on policies, the BDT service SHOULD add security labels to exported resources or MAY update existing security labels.	Export	SLS, PPS
3.2. Security Labeling on Import	Based on policies, the BDT service MAY add security labels to imported resources or update existing security labels.	Import	SLS, PPS
3.3. Client-Based Integrity Labeling	Based on policies, the BDT service SHOULD arrange assigning integrity labels to imported data depending on its level of trust in the client and the origin of the data.	Import	SLS, Provenance Service
3.4. Profile-Based Content Transformation on Export	Based on policies, the BDT service SHOULD apply transformation functions (detailed in standard profiles) on the content of exported resources.	Export	SLS, PPS
3.5. Profile-Based Content Transformation on Import	Based on policies, the BDT service MAY apply transformation functions (detailed in standard profiles) on the content of imported resources.	Import	SLS, PPS
4. Provenance			

Title	Description	Scope	Services
4.1. Generating Origin Provenance for Export	The BDT SHOULD record the circumstances of the export in the form of a provenance resource accompanying the result of the export operation.	Export	Provenance Service
4.2. Consuming Origin Provenance in Import	The BDT SHOULD recognize and consume the provenance information provided by the origin of the data at the time of import.	Import	Provenance Service
4.3. Generating Origin Provenance for Import	The BDT SHOULD record the circumstances of the import in the form of a provenance resource.	Import	Provenance Service
4.4. Transformation and Filtering Provenance	In the event of any transformation or filtering, the BDT MAY record provenance information regarding the transformations applied to the content of resources.	Import, Export	Provenance Service

5 NEXT STEPS

This report highlights the importance of privacy requirements for the BDT and provides an initial draft for such requirements. The following next steps are proposed in order to promote these requirements and to ensure the consideration of privacy requirements in the BDT and FHIR specifications:

- Sharing this report with the HL7 Security Work Group and the FHIR community;
- Collecting feedback from stakeholders about these requirements and other potential privacy requirements for the BDT service, and updating this report based on that input;
- Promoting Privacy-Aware BDT in the future HL7 FHIR Connectathons to engage with the vendors and the developer community; and
- An Implementation Guide for Privacy-Aware BDT either as part of a broader Data Segmentation for Privacy (DS4P) Implementation Guide or as a separate effort.
- Recording these requirements as privacy considerations in the FHIR specifications;

6 REFERENCES

[1] FHIR Bulk Data Access, https://build.fhir.org/ig/HL7/bulk-data/

https://github.com/HL7/bulk-data/blob/master/spec/export/index.md
(fetched on 05/25/2019)

[2] General Data Protection Regulation, Article 22, Automated Individual Decision-Making Including Profiling. https://gdpr-info.eu/art-22-gdpr/

[3] General Data Protection Regulation, Article 35, Data Protection Impact Assessment. https://gdpr-info.eu/art-35-gdpr/

[4] Dutch Data Protection Authority (Dutch DPA), Data Protection Impact Assessment Guidelines, https://autoriteitpersoonsgegevens.nl/nl/zelf-doen/data-protection-impact-assessment-dpia (in Dutch; accessed via machine translation)

7 ACRONYMS

ACS	Access Control Service
API	Application Programming Interface
BDT	Bulk Data Transfer
CBCP	Community-Based Care and Privacy
CDC	Centers for Disease Control and Prevention
CMS	Centers for Medicare, and Medicaid Services
DS4P	Data Segmentation for Privacy
DPIA	Data Protection Impact Assessment
EHR	Electronic Health Record
FHIR	Fast Healthcare Interoperability Resources
GDPR	General Data Protection Regulation
HIPAA	Health Insurance Portability and Accountability Act
HL7	Health Level 7
PPS	Privacy-Preserving Services
SLS	Security Labeling Service
SMART	Substitutable Medical Apps, Reusable Technology