

8.3 Virtual Health Record

A Virtual Health Record (VHR) provides a virtual view of heterogeneous data sources, using a common data model. In contrast to the data warehousing approach in which heterogeneous data is extracted, transformed and stored in a homogeneous form, the VHR approach does not require clinical data to be extracted from existing data stores. Instead, logical queries are defined in terms of a common data model and then transformed into a set of physical queries which can each be executed locally on an individual data store. Figure 8.3-1 illustrates an architecture which supports querying over a VHR.

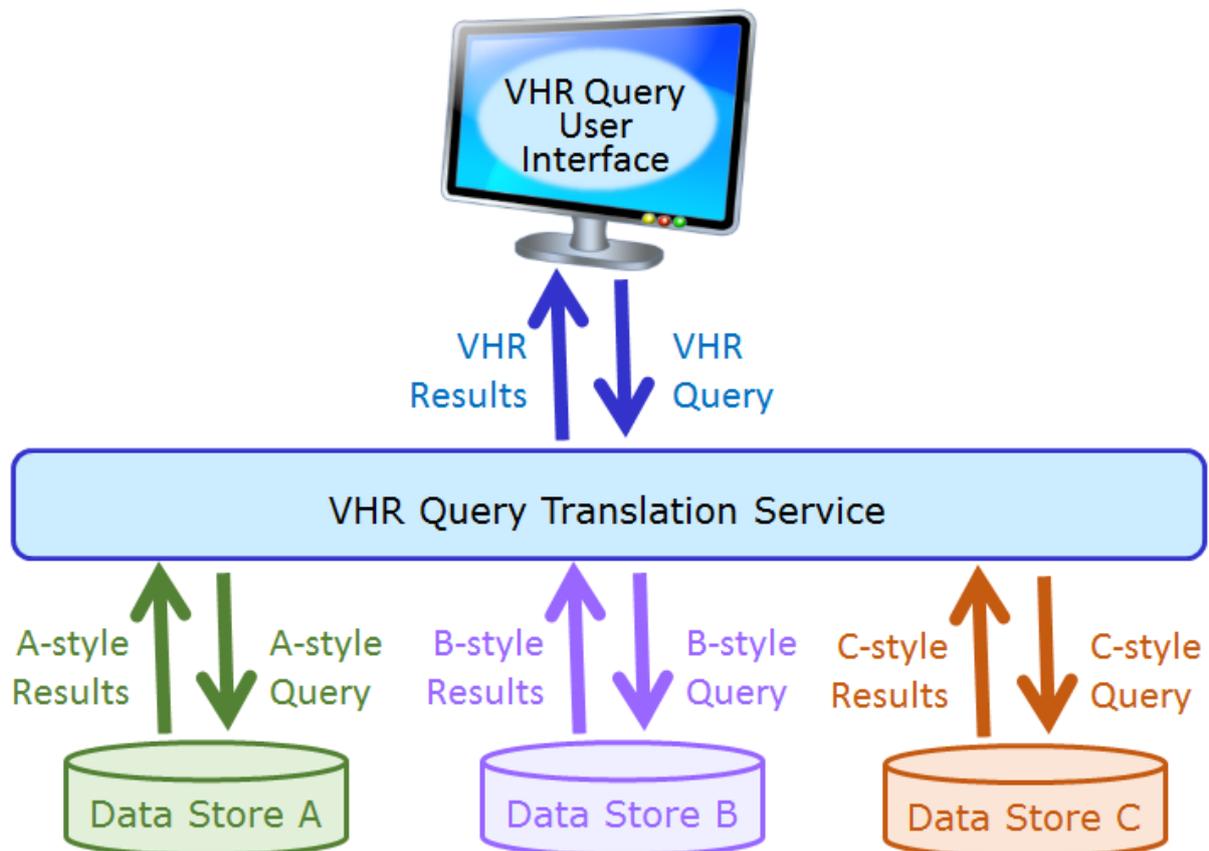


Figure 8.3-1: Querying using a Virtual Health Record

The process of transforming the logical query into separate physical queries may involve translating:

- *The Query Language* – from a common query language to the local data store's native query language
- *Data Model References* – from the common data model to the local data model
- *Terminology References* – from the standard terminology to the local code system

For example, if the user poses the following SQL query, written in terms of the VHR's common data model, to select those patients with a diagnosis that is a subtype of 40733004 |infectious disease|:

```
SELECT patient_id FROM Health_Records  
WHERE diagnosis IN (<40733004 |infectious disease|)
```

This query may be translated into the following 3 queries for local execution on each data store:

Data Store A:

```
Patient_record/patient_id[@diagnosis=typeOf(INF)]
```

Data Store B:

```
SELECT id FROM EHR NATURAL JOIN DSummary  
WHERE discharge_diagnosis IN (descendantsOf (40733004))
```

Data Store C:

```
SELECT patient FROM record
```

```
WHERE diag IN (<40733004)
```

Similarly, when the query results are returned by each data store, these need to be transformed and mapped into the common data model and then combined for presentation to the user.

The VHR approach provides an alternative architecture to a data warehouse for integrating heterogeneous systems. It is most commonly used when copying clinical data into a data warehouse is not possible (e.g. due to legislative requirements), or when the currency of the data is imperative. The challenges with this approach lie with the potential complexity of the transformations required. The implementation of this approach is considered to be a type of heterogeneous distributed database, as described in [Section 8.4 Distributed Storage and Processes](#).
