

Date Filtering on the NBN Gateway

Version History

Version	Date	Description
1	June 2006	Initial draft notes, Richard Ostler

Introduction

The NBN Gateway uses the concept of vague dates to store date information. A vague date specifies a start date and an end date of a date range together with a date type code identifying the form of the date type. This concept provides great flexibility when storing date information allowing anything from an exact date (i.e. 20/08/1976) to a year range (i.e. 1980-1990) or even a pre or post date (i.e. before 1950).

Querying vague dates on the NBN Gateway

Query date granularity

The query date granularity on the NBN Gateway is a year. This means the Gateway will allow you to ask for all records from 1990, but not for the first 6 months of 1990. Therefore, for any date recorded to a day, month, day range or month range only the year part of the vague date is queried against.

Allowed date queries

The NBN Gateway allows users to optionally specify start and end year values for a date range. The start and end year values are inclusive. The table below describes how user entered start and year values are interpreted.

Start year	End Year	Interpretation
1990	1990	All records recorded in 1990
1990	2000	All records recorded between 1990 and 2000 inclusive
1990	blank	All records recorded after and including 1990
blank	2000	All records recorded before and including 2000

Overlapping record year ranges and query year ranges

One consequence of the NBN Gateway recording year ranges and users being able to filter on date ranges is a recorded date range and query date range can overlap (fig 1).

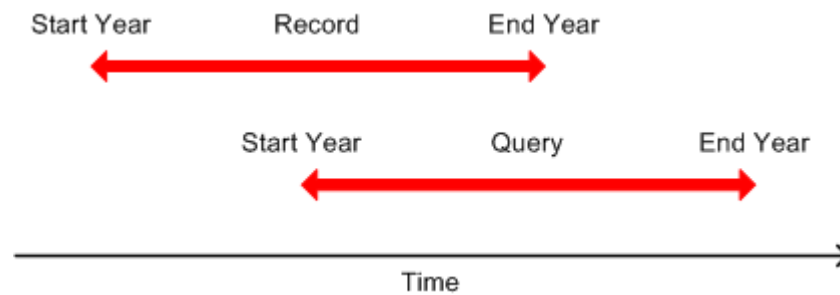


Figure 1.

On the NBN Gateway the design decision has been taken to include any record with a year range if its year range either touches or overlaps, to any magnitude, a user's query range. It is possible a future implementation could give users more control, for example by allowing year range records to be excluded or select records based on some weighting of the degree of overlap between the query year range and recorded year range.

Query Logic

Considering the granularity of the date filtering is year there are three functionally different year range types an NBN record can have:

1. start year \leq end year
2. start year only
3. end year only

start year \leq end year

Here the recorded date range has a known start and end year, for example 1980 to 1990; 2000 to 2000. Figure 2 illustrates the set of 6 possible overlap relations between a recorded year range and query year range.

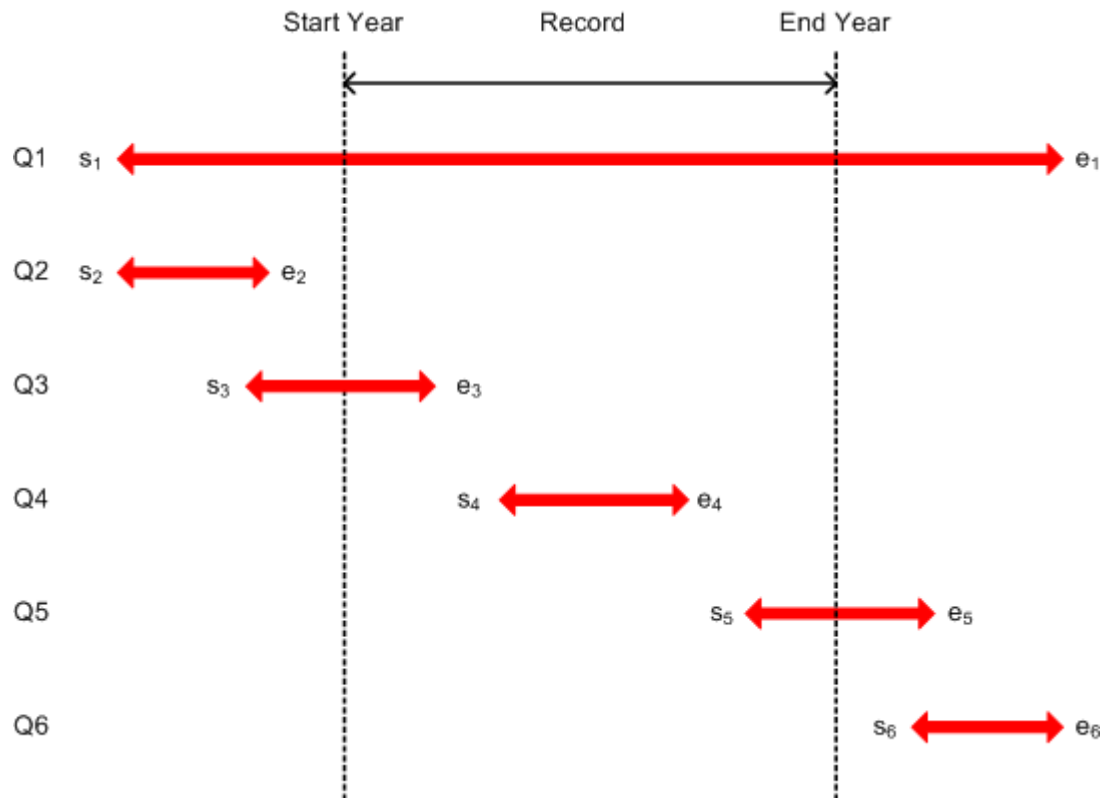


Figure 2.

From figure 2, query date ranges Q1, Q3, Q4 and Q5 have some degree of overlap with the record year range. Therefore those queries should select the record. For query date ranges Q2 and Q6 there is no overlap, so those queries should not select the record. This can be expressed logically as:

query start year \leq record end year AND query end year \geq record start year

start year only

Here the record start year is known, but there is no end year, for example 1970 onwards. Figure 3 illustrates the set of 3 possible overlap relations between a start year only record and query year range.

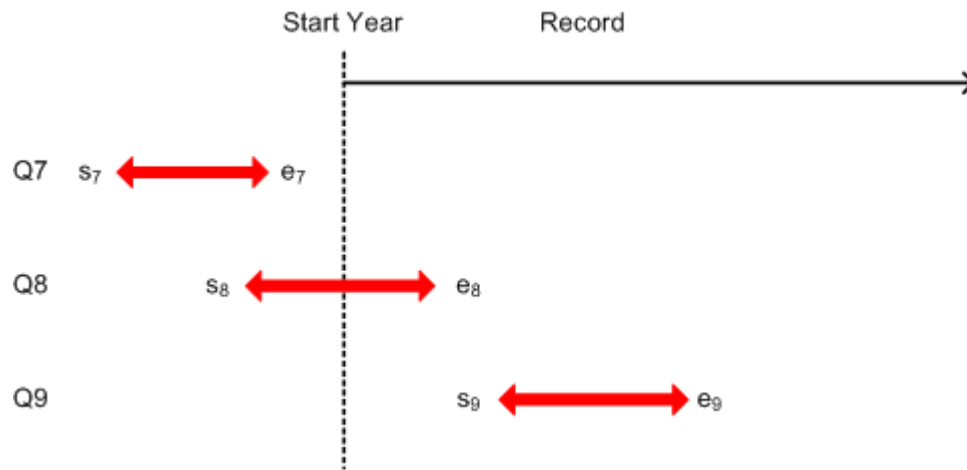


Figure 3.

From figure 3, the query end year values for Q8 and Q9 both fall after the record start year, so the record should be selected. For Q7, there is no overlap so the record would not be selected. This can be expressed logically as:

$$\text{query end year} \geq \text{record start year AND record end year} = 0$$

end year only

Here the record end year is known, but there is no start year, for example before 1970. Figure 4 illustrates the set of 3 possible overlap relations between a end year only record and query year range.

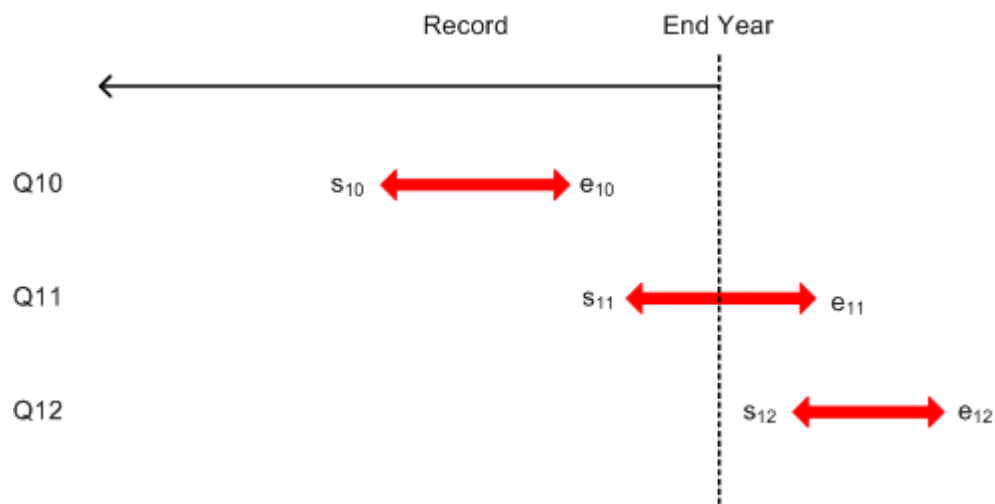


Figure 4.

From figure 4, the query start year values for Q10 and Q11 both fall before the record start year, so the record should be selected. For Q12, there is no overlap so the record would not be selected. This can be expressed logically as:

$$\text{query start year} \leq \text{record end year AND record start year} = 0$$

Implementing query logic in SQL

The Gateway queries should explicitly convert a blank end year to the current year and a blank start year to 0 at runtime. This means all year queries are turned into a date range. This is a design decision which has the following advantages:

1. it negates the need for dynamic SQL queries to be built with conditional predicates based on the year, and other filters
2. reduces the number of coded queries, i.e. you do not need 4 separate queries preidicate (no year filter specified, year range, from year, to year), identical except for their date.

The disadvantage to this is if no year range is specified by a user a year range predicate, which covers all years, is still applied.

The following SQL can be used to implement the query logic:

```
WHERE / AND ((s <= E AND e >= S)
              OR (e >= S AND E = 0)
              OR (s <= E AND S = 0))
```

where: s = query start year
e = query end year
S = record start year
E = record end year