STRING_AGG Function Handout

presented by Joanne Leary to the Reporting Users Group on 11/1/21

Use of the "STRING_AGG" function to concatenate multiple values for a given record (holdings statements, subject headings) into one line of output

The **STRING_AGG** is a very useful way to combine multiple values into a single row of output in a query.

It's very common for serials records (in particular) to have multiple holdings records, and for each of those holdings records to have multiple **holdings statements** (866's) and **holdings notes**. It's also typical for any given work (instance ID) to have multiple subject headings.

- The STRING_AGG function allows you to see all the subject headings and/or holdings statements and/or holdings notes in a single row of a spreadsheet, where each distinct element is separated by an easily recognizable delimiter
- This can be especially useful when comparing holdings of a given instance title across libraries (such as the home library and the Annex), or when you want to search for shared subject terms in a list of titles (expense transfer project)

The components of the STRING_AGG function are: table.field, delimiter

STRING_AGG (table.field, 'delimiter')

- Example: get the statements from the derived table called holdings_statements, which has an alias of "hs", and use a "space_pipe_space" as a delimiter:
  ➔ `STRING_AGG (hs.statements, ' | ')`

- Sometimes there might be repetitions of values (such as in subject headings). To get only the distinct values, enter the word DISTINCT:

  Example: get the distinct subject headings from the instance_subjects derived table:

  ➔ `STRING_AGG (DISTINCT instsubj.subject, ' | ')`

- Aggregation assumes you will be grouping the results on the selected fields in the SELECT section, so you will have a GROUP BY section in your query.

Before and After examples:

Philosophical Transactions. Mathematical, physical and engineering science.

Instance HRID = 2861795

BEFORE:

This is a query to find the holdings records and holdings statements for the title above. We're using the following derived tables:

- Instance_ext (gets the Instance HRID and Title)
- Holdings_ext (gets the Holdings HRID, location, call number and holdings type)
- Holdings_statments (gets the holdings statements)

Without using the aggregation function we have the following SQL:

```
SELECT
	ie.instance_hrid,
	hs.holdings_hrid,
	ie.title,
	he.permanent_location_name,
	he.call_number,
	he.type_name as holdings_type_name,
	hs.statement

FROM
	folio_reporting.holdings_ext as he
		LEFT JOIN folio_reporting.holdings_statements as hs
		ON he.holdings_id = hs.holdings_id

		LEFT JOIN folio_reporting.instance_ext as ie
		ON he.instance_id = ie.instance_id

WHERE
	ie.instance_hrid = '2861795'

ORDER BY holdings_hrid;
```

…which yields the following results:

	inst	holding	title	permanent_location_name	call_number	holding	statement
1	2861795	3386285	Philosophical transactions. Mathematical, pl	Phys Sci - Annex	Q41.L84 A17	Serial	v.354-364 (1996-2006)
2	2861795	3386285	Philosophical transactions. Mathematical, pl	Phys Sci - Annex	Q41.L84 A17	Serial	v.365:no.1851-1861 (2007 Feb.-Dec.)
3	2861795	3386285	Philosophical transactions. Mathematical, pl	Phys Sci - Annex	Q41.L84 A17	Serial	v.366-367 (2008-2009)

AFTER: Now we'll use the STRING_AGG function to combine the holdings statements into one result. We're using the same three tables as in the Before example:

```
SELECT
	ie.instance_hrid,
	hs.holdings_hrid,
	ie.title,
	he.permanent_location_name,
	he.call_number,
	he.type_name AS holdings_type_name,
	string_agg(hs.statement, ' | ') AS holdings_summary

FROM
	folio_reporting.holdings_ext AS he
		LEFT JOIN folio_reporting.holdings_statements AS hs
		ON he.holdings_id = hs.holdings_id

		LEFT JOIN folio_reporting.instance_ext AS ie
		ON he.instance_id = ie.instance_id

WHERE
	ie.instance_hrid = '2861795'

GROUP BY

	ie.instance_hrid,
	hs.holdings_hrid,
	he.permanent_location_name,
	ie.title,
	he.call_number,
	he.type_name

ORDER BY holdings_hrid;
```

Result:

instanc	holdin	title	permanent_locati	call_number	ho	holdings_summary
2861795	3386285	Philosophical transactior	Phys Sci - Annex	Q41.L84 A17	Serial	v.354-364 (1996-2006) | v.365:no.1851-1861 (2007 Feb.-Dec.) | v.366-367 (2008-2009)

FINALLY: You can use more than one STRING_AGG function to get both holdings statements AND subject headings.

```
SELECT
	ie.instance_hrid,
	hs.holdings_hrid,
	ie.title,
	he.permanent_location_name,
	he.call_number,
	he.type_name as holdings_type_name,
	string_agg(hs.statement, ' | ') as holdings_summary,
	string_agg(DISTINCT instsubj.subject, ' | ') as subject_headings

FROM
	folio_reporting.holdings_ext AS he
		left join folio_reporting.holdings_statements AS hs
		on he.holdings_id = hs.holdings_id

		left join folio_reporting.instance_ext AS ie
		on he.instance_id = ie.instance_id

		left join folio_reporting.instance_subjects AS instsubj
		on ie.instance_id = instsubj.instance_id

WHERE
	ie.instance_hrid = '2861795'

GROUP BY
	ie.instance_hrid,
	hs.holdings_hrid,
	he.permanent_location_name,
	ie.title,
	he.call_number,
	he.type_name

ORDER BY holdings_hrid;
```

RESULT: one row, with a column for holdings summary and a column for subject headings

instance_hrid	holdings_hrid	title	permanent_location_name	call_number	holdings_type_name	holdings_summary	subject_headings
2861795	3386285	Philosophical transactio	Phys Sci - Annex	Q41.L84 A17	Serial	v.354-364 (1996-2006) | v.354-364	Engineering | Engineering Periodicals |