May 13, 2021

Presentation on FOLIO LDP MARC: Report from the Metadata Reports Working Group

Jennifer Eustis, *University of Massachusetts Amherst*

Report from the Metadata Reports Working Group

May 13, 2021

In this Presentation ….

- Brief introduction to the MM Reports Working Group
- How to connect to the FOLIO Library Data Platform or LDP
- Introduction to MARC tables in the LDP
- Example reports using MARC data in the LDP

For today, DBeaver will be the tool used for the demos and was used for the screenshots in this presentation.

MM Reports Working Group

The Reporting SIG asked for volunteers from the Metadata SIG to work on testing the LDP (Library Data Platform), developing reports for the LDP, and testing LDP MARC.

For about a year, the reports working group has created a number of derived tables and reports for FOLIO. For the past couple of weeks, we have provided feedback and recommendations for the LDP MARC test.

This working group is open to anyone interested in reports on metadata.

How to Connect to the FOLIO Library Data Platform

To access the database using a client-side reporting tool such as DBeaver:

- Database system: **PostgreSQL**
- Host: **glintcore.net**
- Port: **5432**
- Database name: **folio_snapshot**
- User name and password: (contact Nassib Nassar)
- SSL enabled, SSL mode: **require**

Please note that the password encryption uses scram-sha-256 and may not be compatible with old database tools or drivers.

For more information about using an LDP database:https://github.com/library-data-platform/ldp/blob/main/doc/User_Guide.md

Connection "folio_snapshot" configuration — □ ✕

Connection settings

PostgreSQL connection settings

PostgreSQL

| Connection settings | Main | PostgreSQL | Driver properties | SSH | Proxy | SSL |

- ⌄ Connection settings
  - Initialization
  - Shell Commands
  - Client identification
  - Transactions
- General
- Metadata
- Errors and timeouts
- › Data editor
- › SQL Editor

Server

Host: glintcore.net Port: 5432

Database: folio_snapshot

Authentication

Authentication: Database Native ⌄

Username: ldp

Password: •••••••••••••••• ☑ Save password locally

Advanced

User role:

Local Client: ⌄

ⓘ You can use variables in connection parameters.

Driver name: PostgreSQL   Edit Driver Settings

Test Connection ...   OK   Cancel

File   Edit   Navigate   Search   SQL Editor   Database   Window   Help

SQL

Database Navigator ⊠   Projects

Enter a part of table name here

- folio-snapshot - *glintcore.net:5432*
  - folio_marc_test
  - folio_snapshot
  - folio_snapshot_20201111_feesfines
  - folio_snapshot_20201120_feesfines
  - folio_snapshot_20210114_invoice
  - folio_snapshot_20210218_invoice
  - folio_snapshot_20210406_request
  - folio_snapshot_20210414_circlog
  - folio_snapshot_20210426
  - folio_snapshot_20210427
  - folio_snapshot_20210428
  - folio_snapshot_20210429
  - folio_snapshot_20210430
  - folio_snapshot_20210501
  - folio_snapshot_20210502
  - folio_snapshot_20210503
  - folio_snapshot_20210504
  - folio_snapshot_20210505
  - folio_snapshot_20210506
  - folio_snapshot_20210507
  - folio_snapshot_20210508
  - folio_snapshot_20210509
  - folio_snapshot_20210510
  - folio_snapshot_20210510_voucher
  - folio_snapshot_20210511
  - folio_test_data
  - postgres

Folio Reporting and Public Schemas

- folio_reporting
  - Derived tables: These are helper tables to get at information more easily to create reports.
- public
  - All of the tables from FOLIO including
    - srs_marc

      This is the table that has the marc srs json data.

    - srs_records
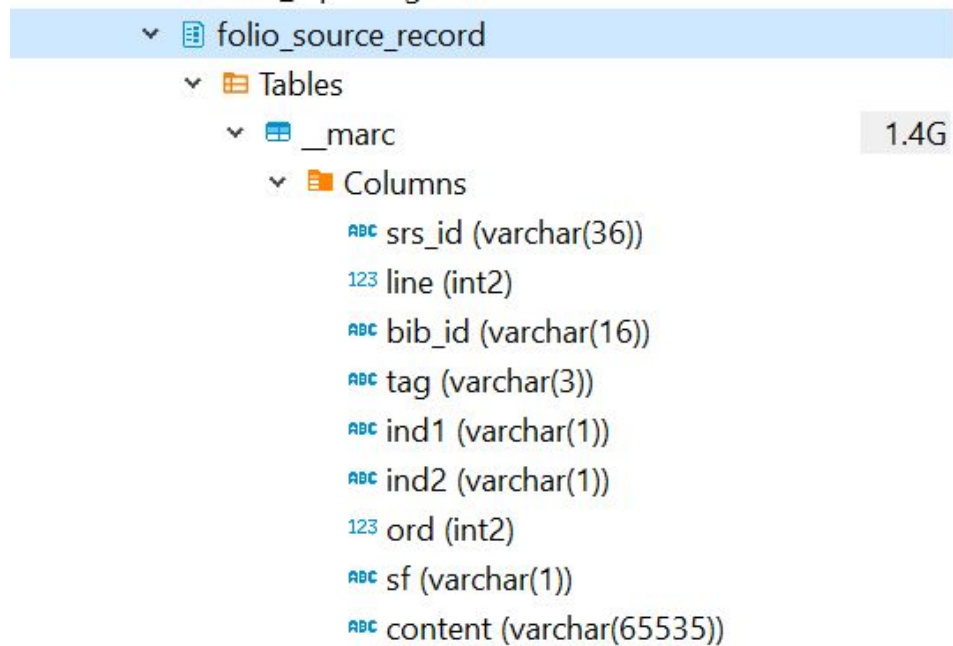
      This is the table that has the administrative metadata and the marc srs json data.
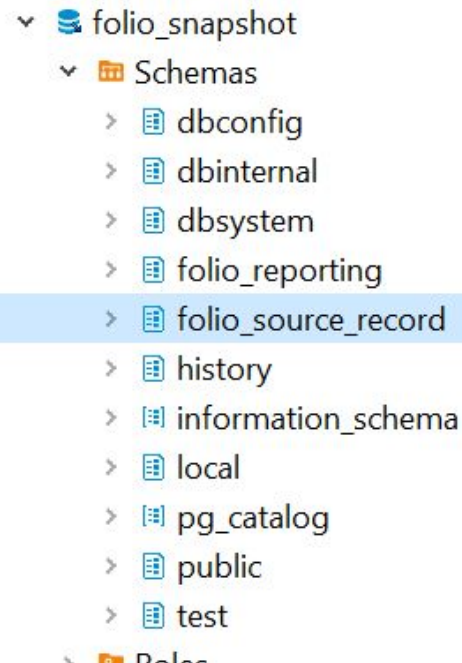
folio_source_record

The __marc table is the "interpreted" or "deconstructed" MARC SRS. This table is still in development and you will see changes. The is test data that sometimes doesn't always make sense. This table includes:

- srs_id: the UUID for the MARC SRS record.
- line: line number for the MARC tag
- bib_id: instance hrid
- tag: MARC tag (field)
- ind1: first indicator
- ind2: second indicator
- ord: ordinality or the order of the tag
- sf: subfield
- content: content in the subfield

What Does This Look Like for SRS_Marc?

id	data	ten
1	997f77c8-672a-11eb-8681-aed9fae510e9	{ "id": "997f77c8-672a-11eb-8681-aed9fae510e9", "fields": [{ "001": "200388703" }, { "003":	1
2	998010ac-672a-11eb-8681-aed9fae510e9	{ "id": "998010ac-672a-11eb-8681-aed9fae510e9", "fields": [{ "001": "200388851" }, { "003":	1
3	99809f90-672a-11eb-8681-aed9fae510e9	{ "id": "99809f90-672a-11eb-8681-aed9fae510e9", "fields": [{ "001": "20038886X" }, { "003":	1
4	99815d72-672a-11eb-8681-aed9fae510e9	{ "id": "99815d72-672a-11eb-8681-aed9fae510e9", "fields": [{ "001": "200388916" }, { "003":	1
5	9982228e-672a-11eb-8681-aed9fae510e9	{ "id": "9982228e-672a-11eb-8681-aed9fae510e9", "fields": [{ "001": "200388924" }, { "003":	1
6	9982d274-672a-11eb-8681-aed9fae510e9	{ "id": "9982d274-672a-11eb-8681-aed9fae510e9", "fields": [{ "001": "200388932" }, { "003":	1
7	9984afd6-672a-11eb-8681-aed9fae510e9	{ "id": "9984afd6-672a-11eb-8681-aed9fae510e9", "fields": [{ "001": "20038984X" }, { "003":	1

What Does This Look Like for SRS_Records?

srs_records   14

∨ 🗀 Columns

- 🔤 id (varchar(36))
- ᴬᴮᶜ created_by_user_id (varchar(36))
- ᴬᴮᶜ created_date (varchar(29))
- 123 generation (int8)
- ᴬᴮᶜ instance_hrid (varchar(13))
- ᴬᴮᶜ instance_id (varchar(36))
- ᴬᴮᶜ leader_record_status (varchar(1))
- ᴬᴮᶜ matched_id (varchar(36))
- 123 order (int8)
- ᴬᴮᶜ record_type (varchar(4))
- ᴬᴮᶜ snapshot_id (varchar(36))
- ᴬᴮᶜ state (varchar(6))
- ☑ suppress_discovery (bool)
- ᴬᴮᶜ updated_by_user_id (varchar(36))
- ᴬᴮᶜ updated_date (varchar(29))
- 🔣 data (json)
- 123 tenant_id (int2)

id	created_by_user_id	created_date	gen
04f09398-672b-11eb-8681-aed9fae510e9	7a16958e-af23-5002-8191-d7a96f5f8070	2021-02-05 20:25:59.158434+00	0
8dbafe52-672b-11eb-8681-aed9fae510e9	7a16958e-af23-5002-8191-d7a96f5f8070	2021-02-05 20:25:59.253498+00	0
c447a53e-672a-11eb-8681-aed9fae510e9	7a16958e-af23-5002-8191-d7a96f5f8070	2021-02-05 20:25:59.449207+00	0
04f40cbc-672b-11eb-8681-aed9fae510e9	7a16958e-af23-5002-8191-d7a96f5f8070	2021-02-05 20:25:59.670102+00	0
c449f2ee-672a-11eb-8681-aed9fae510e9	7a16958e-af23-5002-8191-d7a96f5f8070	2021-02-05 20:25:59.688393+00	0

What Does This Look Like for __marc?

SELECT * FROM folio_marc_test.folio_source_record."__marc" m ;

	srs_id	line	bib_id	tag	ind1	ind2	ord	sf	content
1	3006fc48-672b-11eb-8681-aed9fae510e9	12	010004785	035			2	a	(OCoLC)263588499
2	3006fc48-672b-11eb-8681-aed9fae510e9	13	010004785	040			1	a	9001
3	3006fc48-672b-11eb-8681-aed9fae510e9	14	010004785	040			1	b	ger
4	3006fc48-672b-11eb-8681-aed9fae510e9	15	010004785	040			1	c	DE-101
5	3006fc48-672b-11eb-8681-aed9fae510e9	16	010004785	040			1	d	9999
6	3006fc48-672b-11eb-8681-aed9fae510e9	17	010004785	044			1	c	XA-YUCS
7	3006fc48-672b-11eb-8681-aed9fae510e9	18	010004785	082	7	4	1	a	540
8	3006fc48-672b-11eb-8681-aed9fae510e9	19	010004785	082	7	4	1	a	610
9	3006fc48-672b-11eb-8681-aed9fae510e9	20	010004785	082	7	4	1	q	DE-600

Examples Using the Deconstructed Marc Table

1. How to construct from deconstruction using string agg
2. Getting language where there are no characters from a to z
3. Working with the 008 fixed field and Leader information

Reconstructing the Deconstructed MARC Tag

SELECT

    marc.bib_id,

    marc.tag,

    string_agg('$'::varchar || marc.sf || marc.content, '') AS sf_content

FROM

    folio_marc_test.folio_source_record."__marc" AS marc

WHERE

    marc.tag IN ('336','337','338')

GROUP BY

    marc.bib_id,

    marc.tag

LIMIT 100;

	bib_id	tag	sf_content
1	01000002X	336	$aText$btxt$2rdacontent
2	01000002X	337	$aohne Hilfsmittel zu benutzenbn2rdamedia
3	01000002X	338	$aBand$bnc$2rdacarrier
4	010000038	336	$aText$btxt$2rdacontent
5	010000038	337	$aohne Hilfsmittel zu benutzenbn2rdamedia
6	010000038	338	$aBand$bnc$2rdacarrier
7	010000054	336	$aText$btxt$2rdacontent
8	010000054	337	$aohne Hilfsmittel zu benutzenbn2rdamedia
9	010000054	338	$aBand$bnc$2rdacarrier
10	010000062	336	$aText$btxt$2rdacontent

A Report on Languages from the 008 Fixed Field

SELECT

   m.bib_id,

   substring(m."content", 36,3) AS language

FROM

   folio_marc_test.folio_source_record."__marc" m

WHERE

   m.tag = '008'

AND

   substring(m."content", 36,3) !~ '[a-z]';

	bib_id	language
1	209635665				
2	209636890				
3	209637382				
4	209638443				
5	20963846X	|||
6	209638745				
7	209638788				
8	209638907				
9	209639075				
10	209639199	

Working with the 008 Fixed Field and Leader

```sql
SELECT
        m.bib_id,
        substring(m.CONTENT, 7,2) AS ldr_byte_06,
        substring(m2."content", 7,1) AS date_type,
        ie.mode_of_issuance_name,
        ie.type_name AS instance_type,
        iet.material_type_name AS item_type,
        iet.permanent_location_name
FROM
        folio_marc_test.folio_source_record."__marc" m
        LEFT JOIN folio_reporting.instance_ext AS ie ON m.bib_id = ie.instance_hrid
        LEFT JOIN folio_reporting.item_ext AS iet ON m.bib_id = iet.item_hrid
        LEFT JOIN folio_marc_test.folio_source_record."__marc" AS m2 ON m.bib_id = m2.bib_id
WHERE
        m.tag = '000'
AND
        ie.mode_of_issuance_name != 'serial'
AND
        m2.tag = '008'
LIMIT 100;
```

Working with the 008 Fixed Field

bib_id	ldr_byte_(	date_type	mode_of_	instance_t	item_type	permanent_location_name
209642564	am	s		single unit	text	
209642572	am	s		single unit	text	
209642580	am	s		single unit	text	
209642599	am	s		single unit	text	
209642653	am	s		single unit	text	
209642688	am	s		single unit	text	
209642696	am	s		single unit	text	
209642718	am	s		single unit	text	
209642815	am	s		single unit	text	
209642890	am	s		single unit	text	
209642904	am	s		single unit	text	
209642920	am	s		single unit	text	
209642955	am	s		single unit	text	

The Deconstructed Marc Table is in Development

The working group has drafted a first set of recommendations for how to move forward with the the LDP Marc. As mentioned, this table is in development and is being changed. Don't be surprised to see changes.

Where to find documentation and how to provide feedback?

- MM Reporting Documentation folder
  - Recommendations Document
  - Marc Use Cases and Spreadsheet for metadata reporting priorities
  - Feedback for the LDP Marc test
  - Test reports for the LDP Marc test
- FOLIO-Analytics GitHub
- MARC SRS Administrative Metadata FAQ from Ann-Marie Breaux
- Slack mm-reports-wg

Questions