

Gisaf - Bug #8267

Reconciliation for MM project giving an error of Key (id) already exists

10/05/2019 12:09 - Giulio Di Anastasio

Status:	Rejected	Start date:	10/05/2019
Priority:	Normal	Due date:	
Assignee:	Philippe May	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:		Spent time:	0.00 hour
Description			
sqlalchemy.exc.IntegrityError sqlalchemy.exc.IntegrityError: (psycopg2.IntegrityError) duplicate key value violates unique constraint "RAW_V_TOPO_MSTK_pkey" DETAIL: Key (id)=(2728060639) already exists. [SQL: 'INSERT INTO avsm_raw_survey."RAW_V_TOPO_MSTK" (id, orig_id, date, status, geom, equip_id, srvyr_id, accur_id, project_id) VALUES (s, %(orig_id)s, %(date)s, %(status)s, ST_GeomFromEWKB((ST_GeomFromEWKB_1)s), %(equip_id)s, %(srvyr_id)s, %(accur_id)s, %(project_id)s)'] [parameters: {'id': 2728060639, 'orig_id': '250144', 'date': datetime.date(2017, 11, 20), 'status': 'E', 'ST_GeomFromEWKB_1': <memory at 0x7fbc058dbb88>, 'equip_id': 2, 'srvyr_id': 1, 'accur_id': 1, 'project_id': 1}] (Background on this error at: http://sqlalche.me/e/gkpi)			
Related issues:			
Related to Gisaf - Feature #8721: Survey data basket: handle the case when mo...		New	27/07/2019

History

#1 - 10/05/2019 12:09 - Giulio Di Anastasio

- Project changed from GIS to Gisaf

#2 - 26/07/2019 13:04 - Philippe May

- Status changed from New to Rejected

Investigating.

Found that there are 2 points with the same id 2728060639: one in V_FENC_MESH and one in V_TOPO_ELPT.
It's very unlikely that it can happen (although theoretically possible, that's the role of hash functions).

```
avgis=# select id, orig_id, date, ST_AsText(geom), srvyr_id, equip_id from avsm_raw_survey."RAW_V_TOPO_ELPT" w  
here id=2728060639;
```

id	orig_id	date	st_astext	srvyr_id	equip_id
2728060639	250144	2017-11-20	POINT Z (370524.2535 1327309.99 53.4144)	1	2

(1 row)

```
avgis=# select id, orig_id, date, ST_AsText(geom), srvyr_id, equip_id from avsm_raw_survey."RAW_V_FENC_MESH" w  
here id=2728060639;
```

id	orig_id	date	st_astext	srvyr_id	equip_id
2728060639	50133	2017-11-20	POINT Z (370556.4542 1327308.5898 53.6987)	1	2

(1 row)

I checked in the basket of survey points and found 2 files for that date 2017-11-20 for the same project, equipment and surveyor.

So, the issue isn't about the reconciliation but actually about the workflow: it wasn't designed that the same surveyor and equipment could have several inputs. So, different points end up being assigned the same id (by the hash function), the latest one being imported is actually stored in the raw_survey table:

```
avgis=# select id, date, project_id, orig_id, srvyr_id, category, equip_id from gisaf_survey.raw_survey where  
id = 2728060639;
```

id	date	project_id	orig_id	srvyr_id	category	equip_id
2728060639	2017-11-20	1	250144	1	T15	2

Given that many such duplicate files are in the basket, the issue needs actually a major overview about workflow/process of survey data import.

Closing this ticket: reconciliation works as expected, and we need to open another ticket about the raw survey data import, which should include the mitigation of the data imported so far (possible headache to come...).

#3 - 27/07/2019 12:03 - Philippe May

- *Related to Feature #8721: Survey data basket: handle the case when more than one file in the basket per day added*