

GIS - Bug #10862

Fix the issues resulted from import point bug

30/08/2020 14:49 - Philippe May

Status:	New	Start date:	30/08/2020
Priority:	Normal	Due date:	
Assignee:	Philippe May	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:		Spent time:	0.00 hour
Description			
The bug #10830 caused issues on the map: many points are not where they are supposed to be. Even though #10830 is fixed, this ticket tracks the corrective actions for fixing the database used in production.			
Related issues:			
Related to Gisaf - Bug #10830: Import issue with raw survey points		Resolved	24/08/2020

History

#1 - 30/08/2020 14:49 - Philippe May

- Related to Bug #10830: Import issue with raw survey points added

#2 - 04/09/2020 17:57 - Philippe May

I have written a script to fix automatically the mismatches see (below). It takes the results of the Survey to point mismatch integrity check, and uses the Gisaf code to reimport the files in the basket which were detected by the integrity check, with a special option (remove_misplaced) to fix such an issue.

We'll check the results on the map.

We can also re-run the integrity check.

In case something went wrong, i have verified beforehand that the database backup was OK so we can revert safely.

Results

```
(gisaf_python3.7) gisaf@gisaf2:~$ python Import\ reproject\ #10830.py
INFO:Gisaf registry:Discovered 515 models
/var/local/gisaf/Survey/Archaeological_Site-2017-12-18.csv: 478 points
/var/local/gisaf/Survey/CZ-2020-02-10-TS.txt: 405 points
/var/local/gisaf/Survey/CZ-2020-02-11-TS.txt: 132 points
/var/local/gisaf/Survey/CZ-2020-02-13-TS.txt: 780 points
/var/local/gisaf/Survey/CZ-2020-02-15-TS.txt: 279 points
/var/local/gisaf/Survey/CZ-2020-07-27-TS.csv: 14 points
/var/local/gisaf/Survey/RZ-2020-01-11-RTK.txt: 84 points
/var/local/gisaf/Survey/RZ-2020-01-13-TS.txt: 559 points
/var/local/gisaf/Survey/RZ-2020-01-15-TS.txt: 434 points
/var/local/gisaf/Survey/RZ-2020-01-21-TS.txt: 401 points
/var/local/gisaf/Survey/RZ-2020-01-22-TS.txt: 477 points
/var/local/gisaf/Survey/RZ-2020-01-24-TS.txt: 336 points
/var/local/gisaf/Survey/RZ-2020-02-21-TS.txt: 157 points
/var/local/gisaf/Survey/RZ-2020-02-22-TS.txt: 56 points
/var/local/gisaf/Survey/RZ-2020-02-24-TS.txt: 398 points
/var/local/gisaf/Survey/RZ-2020-02-25-TS.txt: 100 points
/var/local/gisaf/Survey/RZ-2020-03-17-TS.txt: 18 points
/var/local/gisaf/Survey/RZ-2020-03-18-TS.txt: 260 points
/var/local/gisaf/Survey/RZ-2020-03-19-TS.txt: 72 points
/var/local/gisaf/Survey/RZ-2020-03-20-TS.txt: 71 points
/var/local/gisaf/Survey/RZ-2020-05-13-TS.txt: 167 points
/var/local/gisaf/Survey/RZ-2020-05-18-TS.txt: 1150 points
/var/local/gisaf/Survey/RZ-2020-05-25-TS.txt: 814 points
/var/local/gisaf/Survey/RZ-2020-05-29-TS.txt: 207 points
/var/local/gisaf/Survey/RZ-2020-06-01-TS.txt: 349 points
/var/local/gisaf/Survey/RZ-2020-06-02-TS.txt: 373 points
/var/local/gisaf/Survey/RZ-2020-06-03-TS.txt: 193 points
/var/local/gisaf/Survey/RZ-2020-06-04-TS.txt: 182 points
/var/local/gisaf/Survey/RZ-2020-06-08-TS.txt: 565 points
/var/local/gisaf/Survey/RZ-2020-06-09-TS.txt: 294 points
/var/local/gisaf/Survey/RZ-2020-06-10-TS.txt: 282 points
```

```

/var/local/gisaf/Survey/RZ-2020-06-12-TS.txt: 217 points
/var/local/gisaf/Survey/RZ-2020-06-13-TS.txt: 61 points
/var/local/gisaf/Survey/RZ-2020-06-15-TS.txt: 119 points
/var/local/gisaf/Survey/RZ-2020-06-16-TS.txt: 46 points
/var/local/gisaf/Survey/RZ-2020-06-17-TS.txt: 94 points
/var/local/gisaf/Survey/RZ-2020-06-22-TS.txt: 132 points
/var/local/gisaf/Survey/RZ-2020-06-23-TS.txt: 137 points
/var/local/gisaf/Survey/RZ-2020-06-26-TS.txt: 98 points
/var/local/gisaf/Survey/RZ-2020-06-29-TS.txt: 103 points
/var/local/gisaf/Survey/RZ-2020-06-30-TS.txt: 57 points
/var/local/gisaf/Survey/RZ-2020-07-03-TS.txt: 101 points
/var/local/gisaf/Survey/RZ-2020-07-06-TS.txt: 88 points
/var/local/gisaf/Survey/RZ-2020-07-07-TS.txt: 123 points
/var/local/gisaf/Survey/RZ-2020-07-08-TS.txt: 152 points
/var/local/gisaf/Survey/RZ-2020-07-10-TS.txt: 225 points
/var/local/gisaf/Survey/RZ-2020-07-11-TS.txt: 108 points
/var/local/gisaf/Survey/RZ-2020-07-14-TS.txt: 100 points
/var/local/gisaf/Survey/RZ-2020-07-15-TS.txt: 92 points
/var/local/gisaf/Survey/RZ-2020-07-16-TS.txt: 146 points
/var/local/gisaf/Survey/RZ-2020-07-17-TS.txt: 88 points
/var/local/gisaf/Survey/RZ-2020-07-20-TS.txt: 165 points
/var/local/gisaf/Survey/RZ-2020-07-21-TS.txt: 89 points
/var/local/gisaf/Survey/RZ-2020-07-22-TS_Afternoon.txt: 63 points
/var/local/gisaf/Survey/RZ-2020-07-22-TS_Morning.txt: 63 points
/var/local/gisaf/Survey/RZ-2020-07-23-TS.txt: 108 points
/var/local/gisaf/Survey/RZ-2020-07-24-TS.txt: 132 points
/var/local/gisaf/Survey/RZ-2020-07-28-TS.txt: 149 points
/var/local/gisaf/Survey/RZ-2020-07-29-TS.txt: 158 points
/var/local/gisaf/Survey/RZ-2020-07-31-TS.txt: 146 points
/var/local/gisaf/Survey/RZ-2020-08-04-TS.txt: 80 points
/var/local/gisaf/Survey/WATER_PROJECT/MM/AVSM/TS/MM-2020-01-13-TS.txt: 559 points
/var/local/gisaf/Survey/WATER_PROJECT/MM/AVSM/TS/MM-2020-02-15-TS.txt: 279 points

```

Script

```

import numpy as np
import os
import re
import asyncio
from datetime import date
from pathlib import Path

os.environ['USE_PYGEOS'] = '0'
import pandas as pd
import geopandas as gpd

from gisaf.ipynb_tools import Gisaf
from gisaf.config import conf

fname_search_re = re.compile('^(\S+)-(\d\d\d\d)-(\d\d)-(\d\d).*')
def get_date(row):
    match = fname_search_re.match(row['name'])
    if match:
        return date(year=int(match.group(2)), month=int(match.group(3)), day=int(match.group(4)))

async def main():
    ## Instantiate Gisaf module, and discover the models (registry)
    gs = Gisaf()
    await gs.setup(use_pygeos=False)
    await gs.make_models(with_categories=True)

    from gisaf.importers import RawSurveyImporter
    from gisaf.models.admin import FileImport

    miss = await gs.live_server.store.get_gdf('live:Survey to point mismatch')
    miss['ddate'] = miss.date.dt.date
    len(miss)

    importer = RawSurveyImporter()

    ## Get basket files

```

```

file_imports = await FileImport.get_df(where=FileImport.basket=='Survey', with_related=True)
file_imports = file_imports[~file_imports['name'].isna()]
file_imports.rename(columns={
    'gisaf_survey_surveyor_name': 'surveyor',
    'gisaf_survey_equipment_name': 'equipment',
}, inplace=True)

file_imports['date'] = file_imports.apply(get_date, axis=1)

merged_file_imports = file_imports.merge(
    miss,
    left_on=['surveyor_id', 'equipment_id', 'date'],
    right_on=['srvyr_id', 'equip_id', 'ddate'],
    suffixes=('', '_miss'),
)

base_dir = Path(conf.admin['basket']['base_dir'])/'Survey'

#mmiss['path'] = mmiss.apply(lambda row: base_dir/row['name'], axis=1)
merged_file_imports['path'] = merged_file_imports.apply(lambda row: base_dir/row['dir']/row['name'], axis=
1)

for path, fi in merged_file_imports.groupby('path'):
    if not path.exists():
        #print(f'Missing {path}')
        continue
    print(f'{path}: {len(fi)} points')
    result = await importer.do_import(fi.iloc[0], dry_run=False, remove_misplaced=True)
    #print(result.details)

if __name__ == '__main__':
    asyncio.run(main())

```

#3 - 05/09/2020 15:11 - Philippe May

After the integrity check run again, 93 points (in 4 files), were not fixed. Why? Good question, thanks for asking wink.png..

Running the same script fixed the issue:

```

(gisaf_python3.7) gisaf@gisaf2:~$ python Import\ reproject\ #10830.py
INFO:Gisaf registry:Discovered 515 models

```

```
/var/local/gisaf/Survey/CZ-2020-02-14-TS.txt: 20 points  
/var/local/gisaf/Survey/RZ-2019-12-28-TS.txt: 24 points  
/var/local/gisaf/Survey/RZ-2020-02-04-TS.txt: 5 points  
/var/local/gisaf/Survey/RZ-2020-06-25-TS.txt: 44 points
```