

# ETL for MN Geospatial Commons

## Search, collect, & download data

<https://gisdata.mn.gov/content/?q=help/api> (<https://gisdata.mn.gov/content/?q=help/api>)

```
In [3]: # Import sherpa    https://www.pythonsherpa.com/tutorials/2/
import pprint
import requests
import json
import pandas as pd
```

```
In [4]: # Package Lists. MN Geo requires generic auth=('user', 'pass')
response = requests.get('http://gisdata.mn.gov/api/3/action/package_list', aut
h=('user', 'pass'), verify=False)
```

C:\Program Files\ArcGIS\Pro\bin\Python\envs\arcgispro-py3\lib\site-packages\urllib3\connectionpool.py:1004: InsecureRequestWarning: Unverified HTTPS request is being made to host 'gisdata.mn.gov'. Adding certificate verification is strongly advised. See: <https://urllib3.readthedocs.io/en/latest/advanced-usage.html#ssl-warnings>  
InsecureRequestWarning,

```
In [5]: #response = requests.get(packages)

# Use the json module to load CKAN's response into a dictionary
response_dict = json.loads(response.content)

# Check the contents of the response
assert response_dict['success'] is True # make sure if response is OK
```

```
In [6]: datasets = response_dict['result']          # extract all the packages from the
response
print(len(datasets))                             # returns all datasets
```

901

```
In [7]: datasets[-20:] # Display metadata for packages
```

```
Out[7]: ['water-lake-basin-morphology',  
         'water-lake-bathymetry',  
         'water-lake-bathy-shaded-relief',  
         'water-lake-superior-basin',  
         'water-large-lakes-in-minnesota',  
         'water-major-river-centerlines',  
         'water-measured-kittle-routes',  
         'water-mgmt-area-surface-water',  
         'water-mn-public-waters',  
         'water-national-hydrography-data',  
         'water-nat-wetlands-inv-2009-2014',  
         'water-nat-wetlands-inventory',  
         'water-prehist-hydrography',  
         'water-public-drainage-systems',  
         'water-shallow-lakes-id-by-wldlif',  
         'water-strahler-stream-order',  
         'water-trout-streams-pls-sections',  
         'water-wellhead-protection-areas',  
         'water-wild-and-scenic-rec-rivers',  
         'whaf']
```

## Select package & generate URL

```
In [8]: package = 'water-major-river-centerlines' # Define dataset
```

```
In [9]: # Base url for package information.
base_url = 'https://gisdata.mn.gov/api/3/action/package_show?id='

# Construct the url for the package of interest
package_information_url = base_url + package

# Make the HTTP request. MN Geo needs auth=('user', 'pass')
package_information = requests.get(package_information_url, auth=('user', 'pass'), verify=False)

# Use the json module to load CKAN's response into a dictionary
package_dict = json.loads(package_information.content)

# Check the contents of the response.
assert package_dict['success'] is True # again make sure if response is OK
package_dict = package_dict['result'] # we only need the 'result' part from the dictionary

print(package_dict)
```

```
{'license_title': 'License not specified', 'maintainer': None, 'relationships_as_object': [], 'private': False, 'maintainer_email': None, 'num_tags': 1, 'id': 'fa8b1ae1-e8f9-4cc0-87b5-982014dd080a', 'metadata_created': '2014-10-08T13:53:53.211804', 'metadata_modified': '2020-07-09T08:12:33.898447', 'author': None, 'author_email': None, 'state': 'active', 'version': None, 'creator_user_id': None, 'type': 'dataset', 'resources': [{ 'mimetype': None, 'cache_url': None, 'hash': '', 'description': '', 'gdrsResGuid': '', 'cache_last_updated': None, 'url': 'https://resources.gisdata.mn.gov/pub/gdrs/data/pub/us_mn_state_dnr/water_major_river_centerlines/metadata/preview.jpg', 'name': 'Static Preview - Sample Image', 'format': 'JPEG', 'package_id': 'fa8b1ae1-e8f9-4cc0-87b5-982014dd080a', 'created': '2020-07-09T08:12:34.159461', 'state': 'active', 'mimetype_inner': None, 'last_modified': None, 'position': 0, 'revision_id': 'd9830a4b-ec71-401d-ac52-014805b3eccc', 'url_type': None, 'id': '5c2b7a36-0c3c-4b13-8141-f624542bd2cf', 'resource_type': 'preview', 'size': None}, { 'mimetype': None, 'cache_url': None, 'hash': '', 'description': '', 'gdrsResGuid': '{a0413046-4bd5-4719-95ac-6f957b69f0e5}', 'cache_last_updated': None, 'url': 'https://resources.gisdata.mn.gov/pub/gdrs/data/pub/us_mn_state_dnr/water_major_river_centerlines/shp_water_major_river_centerlines.zip', 'name': 'Shapefile', 'format': 'SHP', 'package_id': 'fa8b1ae1-e8f9-4cc0-87b5-982014dd080a', 'created': '2020-07-09T08:12:34.159473', 'state': 'active', 'mimetype_inner': None, 'last_modified': None, 'position': 1, 'revision_id': 'd9830a4b-ec71-401d-ac52-014805b3eccc', 'url_type': None, 'id': '42316257-cda5-45e1-adcc-b24e298d60b8', 'resource_type': 'shp', 'size': None}, { 'mimetype': None, 'cache_url': None, 'hash': '', 'description': '', 'gdrsResGuid': '{43077cc9-8bcb-41ab-9905-dcb3776b7900}', 'cache_last_updated': None, 'url': 'https://resources.gisdata.mn.gov/pub/gdrs/data/pub/us_mn_state_dnr/water_major_river_centerlines/fgdb_water_major_river_centerlines.zip', 'name': 'ESRI File Geodatabase', 'format': 'fgdb', 'package_id': 'fa8b1ae1-e8f9-4cc0-87b5-982014dd080a', 'created': '2020-07-09T08:12:34.159477', 'state': 'active', 'mimetype_inner': None, 'last_modified': None, 'position': 2, 'revision_id': 'd9830a4b-ec71-401d-ac52-014805b3eccc', 'url_type': None, 'id': '02d18c97-da34-4889-88f0-c1f1142dc2d1', 'resource_type': 'fgdb', 'size': None}, { 'mimetype': None, 'cache_url': None, 'hash': '', 'description': '', 'gdrsResGuid': '{f9aed3d7-b903-4539-a45c-f6f69b7b67cd}', 'cache_last_updated': None, 'url': 'https://resources.gisdata.mn.gov/pub/gdrs/data/pub/us_mn_state_dnr/water_major_river_centerlines/gpkg_water_major_river_centerlines.zip', 'name': 'OGC GeoPackage', 'format': 'gpkg', 'package_id': 'fa8b1ae1-e8f9-4cc0-87b5-982014dd080a', 'created': '2020-07-09T08:12:34.159480', 'state': 'active', 'mimetype_inner': None, 'last_modified': None, 'position': 3, 'revision_id': 'd9830a4b-ec71-401d-ac52-014805b3eccc', 'url_type': None, 'id': '9dcca4fc-6bbe-48d4-afb9-b3ed28228c3c', 'resource_type': 'gpkg', 'size': None}, { 'mimetype': None, 'cache_url': None, 'hash': '', 'description': '', 'gdrsResGuid': 'None', 'cache_last_updated': None, 'url': 'https://resources.gisdata.mn.gov/pub/gdrs/data/pub/us_mn_state_dnr/water_major_river_centerlines/metadata/metadata.html', 'name': 'Full Metadata Record', 'format': 'HTML', 'package_id': 'fa8b1ae1-e8f9-4cc0-87b5-982014dd080a', 'created': '2020-07-09T08:12:34.159485', 'state': 'active', 'mimetype_inner': None, 'last_modified': None, 'position': 4, 'revision_id': 'd9830a4b-ec71-401d-ac52-014805b3eccc', 'url_type': None, 'id': 'a9679218-b89f-4e0d-a54c-eeee009b3381', 'resource_type': 'metadata', 'size': None}], 'num_resources': 5, 'tags': [{ 'vocabulary_id': None, 'state': 'active', 'display_name': 'inland waters', 'id': '141554ed-e5fe-4645-b64f-1c62bb66c4ab', 'name': 'inland waters'}], 'groups': [{ 'display_name': 'Inland Waters', 'description': 'Inland water features, drainage systems and their characteristics. Examples: rivers and glaciers, salt lakes, water utilization plans, dams, currents, floods, water quality, hydrographic charts', 'image_display_url': 'https://gisdata.mn.gov/categoryIcons/inlandwaters.png', 'title': 'Inland Waters', 'id': '2204b2e5-7990-4671-94cb-c94c161875df', 'name': 'inland-waters'}], 'license_id': 'notspecified'}
```

```
d', 'relationships_as_subject': [], 'organization': {'description': "Working with citizens to conserve and manage the state's natural resources.", 'created': '2013-11-27T08:45:38.519857', 'title': 'Natural Resources Department', 'name': 'us-mn-state-dnr', 'is_organization': True, 'state': 'active', 'image_url': 'https://gisdata.mn.gov/agencyLogos/mnLogo.jpg', 'revision_id': 'd4bec708-0afd-4421-a810-89383f5b1786', 'type': 'organization', 'id': '0c5bf59b-2a39-402d-bf12-bbf11cf3f201', 'approval_status': 'approved'}, 'name': 'water-major-river-centerlines', 'isopen': False, 'url': None, 'notes': 'This data layer represents stream centerlines for major rivers in Minnesota.<br/><br/>Major River Centerlines are maintained as tabular data and displayed as linear events on the <a href="https://resources.gisdata.mn.gov/pub/gdrs/data/pub/us_mn_state_dnr/water_major_river_centerlines/metadata/../../water_measured_kittle_routes/metadata/metadata.html" target=_blank><b>Stream Routes with Kittle Numbers and Mile Measures</b></a> layer. In the attribute table, designated segments extend from the FROM_MEAS (mile) to the TO_MEAS (mile) and have a total length = [LENGTH_MI] on a route with total length = [ROUTE_MI].', 'owner_org': '0c5bf59b-2a39-402d-bf12-bbf11cf3f201', 'extras': [{'key': 'dsMetadataUrl', 'value': 'https://resources.gisdata.mn.gov/pub/gdrs/data/pub/us_mn_state_dnr/water_major_river_centerlines/metadata/metadata.html'}, {'key': 'dsModifiedDate', 'value': '2018-11-16 00:35:13'}, {'key': 'dsOriginator', 'value': 'Minnesota Department of Natural Resources (DNR)'}, {'key': 'dsPeriodOfContent', 'value': '10/30/2013'}, {'key': 'dsPurpose', 'value': 'Cartographic purposes only.'}, {'key': 'gdrsDsGuid', 'value': '{a4d19b9b-7972-4162-b7bb-076977206f64}'}, {'key': 'spatial', 'value': '{"type": "Polygon", "coordinates": [[[-97.252252, 43.098198], [-97.252252, 49.340250], [-89.406737, 49.340250], [-89.406737, 43.098198], [-97.252252, 43.098198]]]}'}, 'title': 'Stream Routes - Major River Centerlines', 'revision_id': '3707455e-57c0-4a06-9780-ea8c94cd7945'}
```

C:\Program Files\ArcGIS\Pro\bin\Python\envs\arcgispro-py3\lib\site-packages\urllib3\connectionpool.py:1004: InsecureRequestWarning: Unverified HTTPS request is being made to host 'gisdata.mn.gov'. Adding certificate verification is strongly advised. See: <https://urllib3.readthedocs.io/en/latest/advanced-usage.html#ssl-warnings>

InsecureRequestWarning,

```
In [10]: # Get the url for the data from the dictionary
data_url = package_dict['resources'][3]['url'] # [3] access URL within metadata
print('Data url:      ' + data_url)
```

Data url: https://resources.gisdata.mn.gov/pub/gdrs/data/pub/us\_mn\_state\_dnr/water\_major\_river\_centerlines/gpkg\_water\_major\_river\_centerlines.zip

## Download zipfile

```
In [48]: import requests
url = 'https://resources.gisdata.mn.gov/pub/gdrs/data/pub/us_mn_state_dnr/water_major_river_centerlines/gpkg_water_major_river_centerlines.zip'

# download the file contents in binary format
r = requests.get(url)

with open(r"C:\Users\mmMary\Documents\GIS_Classes\GIS_5572\Labs\Lab1\Test_files_download\rivers.zip", "wb") as zip:
    zip.write(r.content)
```