

第7回 PostgreSQL アンカンファレンス@東京 (2016528)

OGR_FDW で CSV もエクセルも!

- 外部テーブル + Import Foreign Schema (9.5+) + OGR_FDW
- フォルダ内の複数 CSV を、一括・簡単に外部テーブル化
- エクセルファイル (xls, xlsx) も外部テーブル化

Satoshi Koda satkouda@gmail.com

このスライド → 来週 UP 予定 <http://kenpg.bitbucket.org/>

「OGR_FDW 使う」とかで検索して下さい

2月20日にも紹介しましたが (第2回 MyNA・JPUG 合同DB勉強会) その詳細 & 追加版です

- 外部テーブル? FDW? という方へ
- 9.5 の新機能 Import Foreign Schema
- OGR_FDW とは
- 日本語の实在データ（国勢調査 csv, xls）で検証
- 嵌まりがちな所、その対策

外部テーブル? FDW? という方へ

- 分かりやすい記事：サイオス OSS よろずブログ 稲垣さん
「 PostgreSQL の他データベースにアクセスする仕組み（FDW 編）」
<http://sios-oss.blogspot.jp/2014/05/postgresqlfdwforeign-data-wrapper.html>
- FDW は仕組み（自 DB 以外の様々なデータソースにアクセス）
Foreign Server と Foreign Table：仕組みを構成するオブジェクト
- 実際：○○○_fdw という Extension をインストールし、
それに基く Foreign Server と Foreign Table を作る
- 操作クエリ：どの FDW でも共通（ラッパーだから）
Foreign Table への SELECT クエリ：基本、普通のテーブルと同じ

いろんなFDWがある

- file_fdw (9.1+、 contrib)
<https://www.postgresql.jp/document/9.5/html/file-fdw.html>
- postgres_fdw(9.3+、 contrib、 花田さん作)
<https://www.postgresql.jp/document/9.5/html/postgres-fdw.html>
- www_fdw(9.2+、 サードパーティ)
https://github.com/cyga/www_fdw
- その他たくさん -> <https://wiki.postgresql.org/wiki/FDW>

- What links here
- Related changes
- Special pages
- Printable version
- Permanent link

Generic SQL Database Wrappers

Data Source	Type	Licence	Code	Install	Doc	Notes
ODBC	Native		github	PGXN	example	Does not compile with PostgreSQL >= 9.2!
ODBC	Native		github			A patched, but completely untested version
JDBC	Native		github			Not maintained ?
JDBC2	Native		github			
SQLAlchemy	Multicorn	PostgreSQL	GitHub	PGXN	documentation	Can be used to access data stored in any database supported by the sqlalchemy
VirtDB	Native	GPL	GitHub		short intro	A generic FDW to access VirtDB data sources (SAP ERP, Oracle)

Specific SQL Database Wrappers

Data Source	Type	Licence	Code	Install	Doc	Notes
PostgreSQL	Native	PostgreSQL	git.postgresql.org		documentation	
Oracle	Native	PostgreSQL	github	PGXN	website	
MySQL	Native		github	PGXN	example	An early version of the Foreign Data Wrapper for MySQL that s
MySQL	Native		github	PGXN	example	An updated FDW for MySQL that support Postgr
Informix	Native	PostgreSQL	github			
Firebird	Native		github	PGXN		currently work-in-progress.
SQLite	Native		github			An FDW for SQLite3 (read-on
Sybase / MS SQL Server	Native		github	PGXN		An FDW for Sybase and Microsoft SQL
MonetDB	Native		github			

NoSQL Database Wrappers

Data Source	Type	Licence	Code	Install	Doc	Notes
Cassandra2	Native		Github			
Cassandra	Multicorn	PostgreSQL	Github			
CouchDB	Native	PostgreSQL	Github	PGXN		Original version
CouchDB	Native	PostgreSQL	Github			golgauth version (9.1 - 9.2+ compatible)
Kyoto Tycoon	Native	MIT	Github			
MongoDB	Native	GPL3+	Github	PGXN	README	EDB version
MongoDB	Multicorn		Github			

File Wrappers

Data Source	Type	Licence	Code	Install	Doc	Notes
CSV	Native	PostgreSQL			documentation	Delivered as an official extension of PostgreSQL 9.1 / example / Another example
CSV	Multicorn	PostgreSQL	GitHub	PGXN	documentation	Each column defined in the table will be mapped, in order, against columns in the table
CSV / Text Array	Native		GitHub		How to	Another CSV wrapper
CSV / Fixed-length	Native		GitHub			
CSV / gzipped	Multicorn		GitHub			PostgreSQL Foreign Data Wrapper for gzipped cvs file
Compressed File	Native		GitHub			
Document Collection	Native	PostgreSQL	GitHub		wiki	
JSON	Native	GPL3	GitHub		Example	
Multi-File	Multicorn	PostgreSQL	GitHub	PGXN	doc	Access data stored in various files in a filesystem. The files are looked up based on a pattern, and parts of the files are mapped to columns, as well as the file's content itself.
Multi CDR	Native	PostgreSQL	GitHub	PGXN		
pg_dump	Native	New BSD	GitHub			Allows querying of data directly against Postgres custom format files created by pg_dump
TAR Files	Native		GitHub			
XML	Multicorn	PostgreSQL	GitHub	PGXN		
ZIP Files	Native		GitHub			

Geo Wrappers

Data Source	Type	Licence	Code	Install	Doc	Notes
GDAL/OGR	Native		GitHub			A wrapper for data sources with a GDAL/OGR driver, including databases like Oracle, Informix, SQLite as well as file formats. See more .
Geocode / GeoJSON	Multicorn		GitHub			a collection of PostGIS-related foreign data wrappers
Open Street Map PBF	Native		GitHub			

LDAP Wrappers

Data Source	Type	Licence	Code	Install	Doc	Notes
LDAP	Native		GitHub	PGXN		Allows to query an LDAP server and retrieve data from some pre-configured Organizational Unit

File Wrappers

Data Source	Type	Licence	Code	Install	Doc	Notes
CSV	Native	PostgreSQL			documentation	Delivered as an official extension of PostgreSQL 9.1 / example / Another example
CSV	Multicorn	PostgreSQL	GitHub	PGXN	documentation	Each column defined in the table will be mapped, in order, against columns in the table
CSV / Text Array	Native		GitHub		How to	Another CSV wrapper

基本の流れ

1. ○○○_fdw をシステムにインストール
2.

```
create extension ○○○_fdw; -- DB にインストール
```
3.

```
create server ○●◎  
foreign data wrapper ○○○_fdw options (...);
```
4.

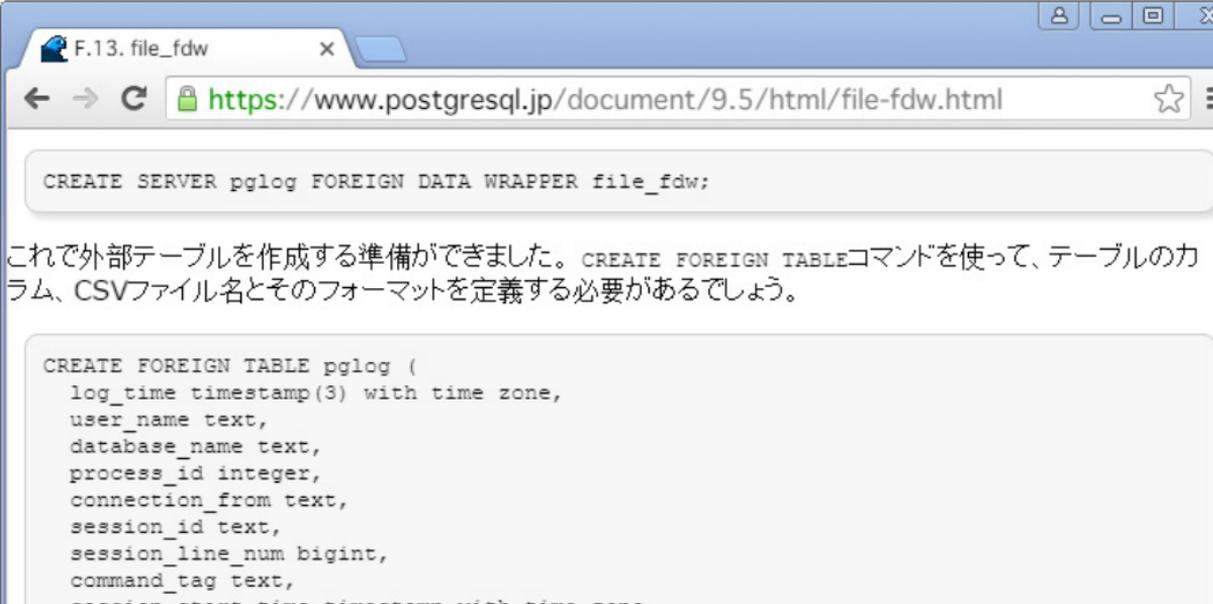
```
create user mapping for ... -- FDW によっては不要  
server ○●◎ options (...);
```
5.

```
create foreign table △▲▽ ( ... )  
server ○●◎;
```

面倒なのは、外部テーブルの定義

- ドキュメントにある file_fdw の例 F.1 (一つの CSV)

<https://www.postgresql.jp/document/9.5/html/file-fdw.html>



```
CREATE SERVER pglog FOREIGN DATA WRAPPER file_fdw;
```

これで外部テーブルを作成する準備ができました。CREATE FOREIGN TABLEコマンドを使って、テーブルのカラム、CSVファイル名とそのフォーマットを定義する必要があるでしょう。

```
CREATE FOREIGN TABLE pglog (  
  log_time timestamp(3) with time zone,  
  user_name text,  
  database_name text,  
  process_id integer,  
  connection_from text,  
  session_id text,  
  session_line_num bigint,  
  command_tag text,  
  session_start_time timestamp with time zone
```



```
CREATE SERVER pglog FOREIGN DATA WRAPPER file_fdw;
```

これで外部テーブルを作成する準備ができました。CREATE FOREIGN TABLEコマンドをラム、CSVファイル名とそのフォーマットを定義する必要があります。

```
CREATE FOREIGN TABLE pglog (  
  log_time timestamp(3) with time zone,  
  user_name text,  
  database_name text,  
  process_id integer,  
  connection_from text,  
  session_id text,  
  session_line_num bigint,  
  command_tag text,  
  session_start_time timestamp with time zone,  
  virtual_transaction_id text,  
  transaction_id bigint,  
  error_severity text,  
  sql_state_code text,
```

9.5 の新機能 Import Foreign Schema

- <https://www.postgresql.jp/document/9.5/html/sql-importforeignschema.html>
- 外部 DB のスキーマにある全テーブルを、一括して外部テーブルに
- いちいちテーブル定義する必要がなくなった ^o^

```
import foreign schema リモートスキーマ名  
***  
from server FDW サーバ名  
into ローカルスキーマ名;
```

★★★で、特定テーブルへの絞り込み/除外も可 (limit to, except)

「スキーマ」のないデータソースは無関係？

- 基本的には、そう。例えば file_fdw:
CSV を扱えるが、1 ファイルごとにテーブル定義する必要あり
- ところが、`OGR_FD` は、仮想的スキーマ名 `ogr_all` を導入することで、Import Foreign schema に対応
- 例えば Geodatabase というファイル形式:
複数の GIS レイヤを持つ → 一度に複数の外部テーブルにできる
- さらに、CSV を含むフォルダをデータソースに指定すると、
フォルダ=スキーマとして扱える ^o^

OGR_FD の文書にはないが、実験したらできた

OGR_FDW とは

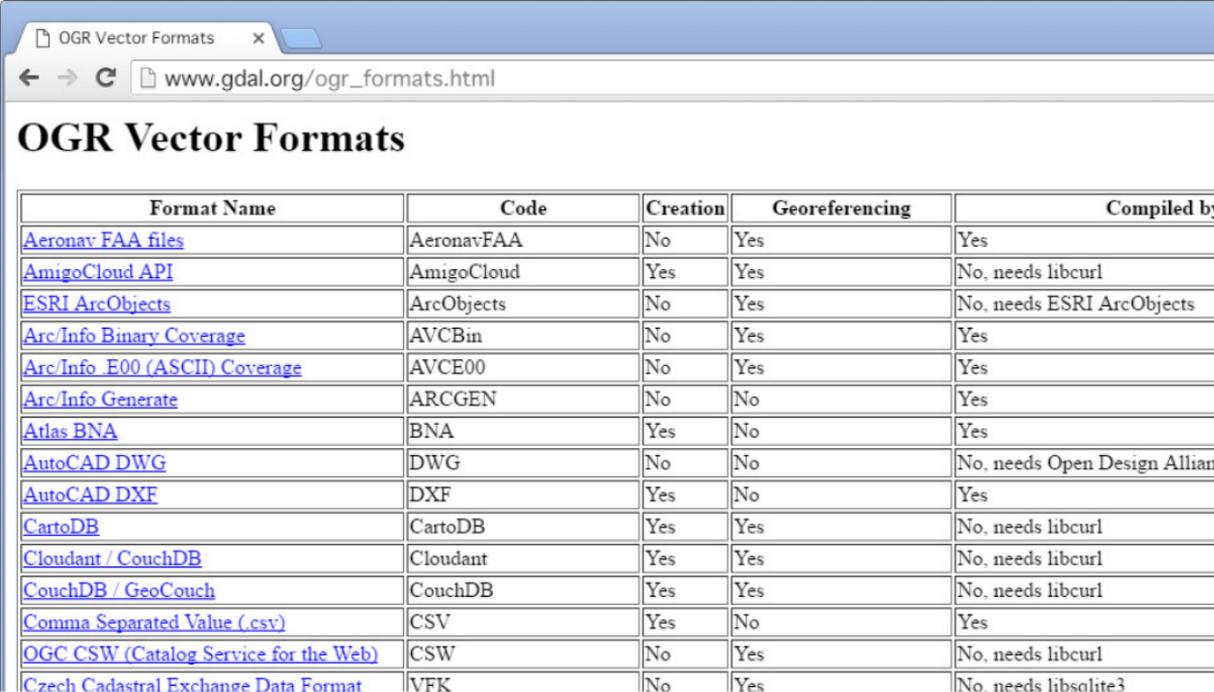
- <https://github.com/pramsey/pgsql-ogr-fdw>
- 作者 … Paul Ramsey ("Mr. PostGIS")
- 主眼 … 多種多様な GIS データを外部テーブル化すること
- でも、GIS と関係ない普通の CSV, Excel も扱える
- 単独で Postgres にインストールできる (PostGIS に依存してない)
- OGR とは … GDAL という GIS 用の定番ライブラリ
に入ってるコマンドラインツール

OpenGIS Simple Features Reference Implementation



OGRが扱えるデータ形式

http://www.gdal.org/ogr_formats.html (現時点で 86 種類)



Format Name	Code	Creation	Georeferencing	Compiled by
Aeronav FAA files	AeronavFAA	No	Yes	Yes
AmigoCloud API	AmigoCloud	Yes	Yes	No, needs libcurl
ESRI ArcObjects	ArcObjects	No	Yes	No, needs ESRI ArcObjects
Arc/Info Binary Coverage	AVCBin	No	Yes	Yes
Arc/Info E00 (ASCII) Coverage	AVCE00	No	Yes	Yes
Arc/Info Generate	ARCGEN	No	No	Yes
Atlas BNA	BNA	Yes	No	Yes
AutoCAD DWG	DWG	No	No	No, needs Open Design Allian
AutoCAD DXF	DXF	Yes	No	Yes
CartoDB	CartoDB	Yes	Yes	No, needs libcurl
Cloudant / CouchDB	Cloudant	Yes	Yes	No, needs libcurl
CouchDB / GeoCouch	CouchDB	Yes	Yes	No, needs libcurl
Comma Separated Value (.csv)	CSV	Yes	No	Yes
OGC CSW (Catalog Service for the Web)	CSW	No	Yes	No, needs libcurl
Czech Cadastral Exchange Data Format	VFK	No	Yes	No, needs libsqlite3

OGRが扱えるデータ形式

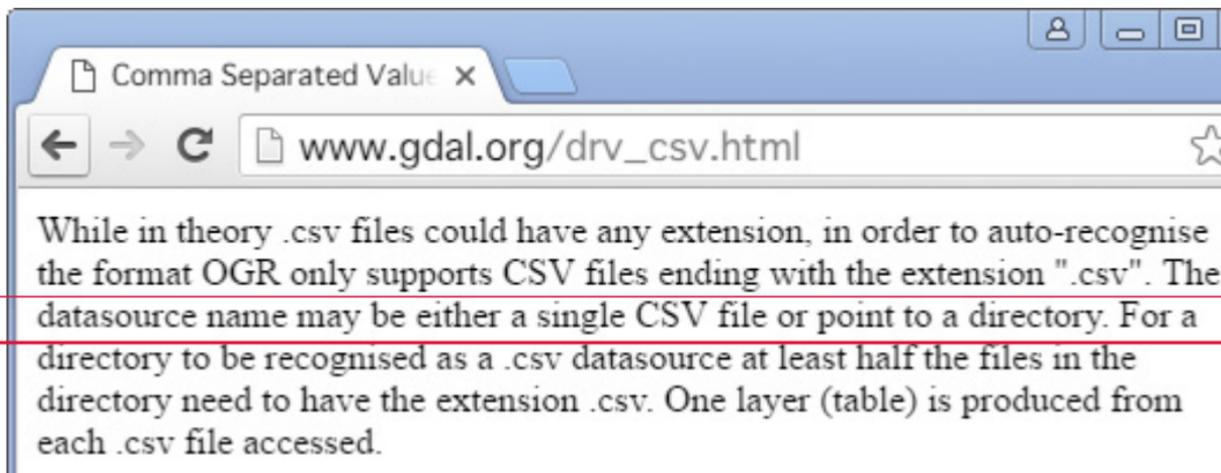
http://www.gdal.org/ogr_formats.html (現時点で 86 種類)

CouchDB / GeoCouch	CouchDB
Comma Separated Value (.csv)	CSV
OGC CSW (Catalog Service for the Web)	CSW
Czech Cadastral Exchange Data Format	VFK
DB2 Spatial	DB2ODBC

VRT - Virtual Datasource	VRT
OGC WFS (Web Feature Service)	WFS
MS Excel format	XLS
MS Office Open XML spreadsheet	XLSX
X-Plane/Flightgear aeronautical data	XPLANE

OGR 文書の CSV 説明の中に…

http://www.gdal.org/drv_csv.html



CSV ファイルだけじゃなく、ディレクトリを渡せるらしいぞ!

で実験したら出来ちゃったという話。

OGR_FDW のインストール

バージョン : 1.0.1 以降を使う (Import Foreign Schema 対応)

rpm で入れるのが簡単 (既に 9.6 用まであり)

[FAQ](#)

[VARIOUS](#)

[BLOG](#)

[DONATE](#)

G+1

159

[YUM REPOSITORY](#)

RPM
1 2 3

Search results for **ogr_fdw** :

Filename	Distribution	File size
ogr_fdw96-1.0.1-1.rhel6.src.rpm	RedHat EL 6	220 k
ogr_fdw96-1.0.1-1.rhel6.src.rpm	RedHat EL 6	220 k
ogr_fdw96-1.0.1-1.f23.src.rpm	Fedora 23	224 k
ogr_fdw96-1.0.1-1.f22.src.rpm	Fedora 22	224 k
ogr_fdw95-debuginfo-1.0.1-1.rhel7.x86_64.rpm	RedHat EL 7	74 k
ogr_fdw95-debuginfo-1.0.1-1.rhel6.x86_64.rpm	RedHat EL 6	74 k
ogr_fdw95-debuginfo-1.0.1-1.rhel6.i686.rpm	RedHat EL 6	74 k
ogr_fdw95-debuginfo-1.0.1-1.f24.x86_64.rpm	Fedora Other	85 k
ogr_fdw95-debuginfo-1.0.1-1.f23.x86_64.rpm	Fedora 23	79 k
ogr_fdw95-debuginfo-1.0.1-1.f22.x86_64.rpm	Fedora 22	78 k
ogr_fdw95-1.0.1-1.rhel7.x86_64.rpm	RedHat EL 7	24 k
ogr_fdw95-1.0.1-1.rhel7.src.rpm	RedHat EL 7	220 k
ogr_fdw95-1.0.1-1.rhel6.x86_64.rpm	RedHat EL 6	21 k

OGR_FDW のインストール for Windows

同様にバージョン 1.0.1 以降を使う（Import Foreign Schema 対応）

コンパイル済みバイナリが PostGIS のウェブサイトにある ↓

<http://winnie.postgis.net/download/windows/pg95/buildbot/extras/>

[\[To Parent Directory\]](#)

Sunday, April 3, 2016 5:49 PM	<dir>	archive
Saturday, April 9, 2016 11:00 PM	6785443	ogrfdw-pg95-binaries-1.0.1w32gcc
Sunday, April 3, 2016 5:52 PM	6614246	ogrfdw-pg95-binaries-1.0.1w64gcc
Saturday, May 7, 2016 2:24 AM	8101377	ogrfdw-pg95-binaries-1.0w32gcc48
Saturday, May 7, 2016 2:23 AM	7901975	ogrfdw-pg95-binaries-1.0w64gcc48
Friday, January 8, 2016 12:49 AM	96068	pointcloud-pg95-binaries-1.1.0de
Friday, January 8, 2016 12:48 AM	61538	pointcloud-pg95-binaries-1.1.0de

ただランタイム libgcc_s_seh-1.dll, libstdc++-6.dll が入っていない。

別途 PostGIS のバイナリなどから持ってくる

以下、実際の経過

実行環境と準備

- Windows 7 64bit + PostgreSQL 9.5.3 + ogr_fdw 1.0.1
- クライアント : ConEmu + Cygwin 版 psql
- 作業用データベースで、ogr_fdw を使う準備

```
# create extension ogr_fdw;  
CREATE EXTENSION
```

-- ランタイムが不足していると ↓ こんなエラーになる

```
ERROR: could not load library "D:/AppsPortable  
/PostgreSQL/9.5.3/lib/ogr_fdw.dll": The specified  
module could not be found.
```

まず foreign server を作る

- スキーマと見なすディレクトリを決め、
- 適当な CSV を一つ以上入れて、
- 下記のようなクエリを実行。

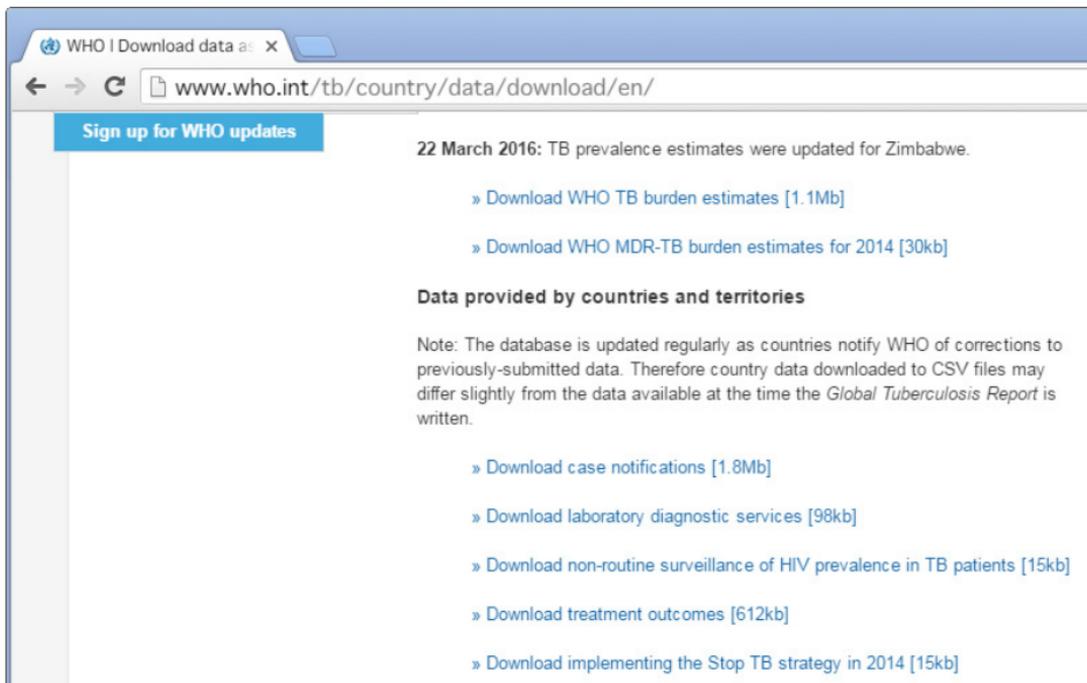
```
create server csv_dir
foreign data wrapper ogr_fdw
options (datasource 'r:/csv_dir', format 'csv');
```

空っぽのディレクトリだと、失敗する。

最低限、カンマ一つのファイルに、拡張子 csv を付けて置けば OK

英語だけのノーマルなCSVからテスト

http://www.who.int/tb/country/data/download/en/



The screenshot shows a web browser window with the address bar displaying "www.who.int/tb/country/data/download/en/". The page content includes a blue button labeled "Sign up for WHO updates". Below this, a date announcement reads "22 March 2016: TB prevalence estimates were updated for Zimbabwe." followed by two links: "» Download WHO TB burden estimates [1.1Mb]" and "» Download WHO MDR-TB burden estimates for 2014 [30kb]". A section titled "Data provided by countries and territories" contains a note about data updates and five links: "» Download case notifications [1.8Mb]", "» Download laboratory diagnostic services [98kb]", "» Download non-routine surveillance of HIV prevalence in TB patients [15kb]", "» Download treatment outcomes [612kb]", and "» Download implementing the Stop TB strategy in 2014 [15kb]".

WHO | Download data as: x

← → ↻ 📄 www.who.int/tb/country/data/download/en/

Sign up for WHO updates

22 March 2016: TB prevalence estimates were updated for Zimbabwe.

- » [Download WHO TB burden estimates \[1.1Mb\]](#)
- » [Download WHO MDR-TB burden estimates for 2014 \[30kb\]](#)

Data provided by countries and territories

Note: The database is updated regularly as countries notify WHO of corrections to previously-submitted data. Therefore country data downloaded to CSV files may differ slightly from the data available at the time the *Global Tuberculosis Report* is written.

- » [Download case notifications \[1.8Mb\]](#)
- » [Download laboratory diagnostic services \[98kb\]](#)
- » [Download non-routine surveillance of HIV prevalence in TB patients \[15kb\]](#)
- » [Download treatment outcomes \[612kb\]](#)
- » [Download implementing the Stop TB strategy in 2014 \[15kb\]](#)

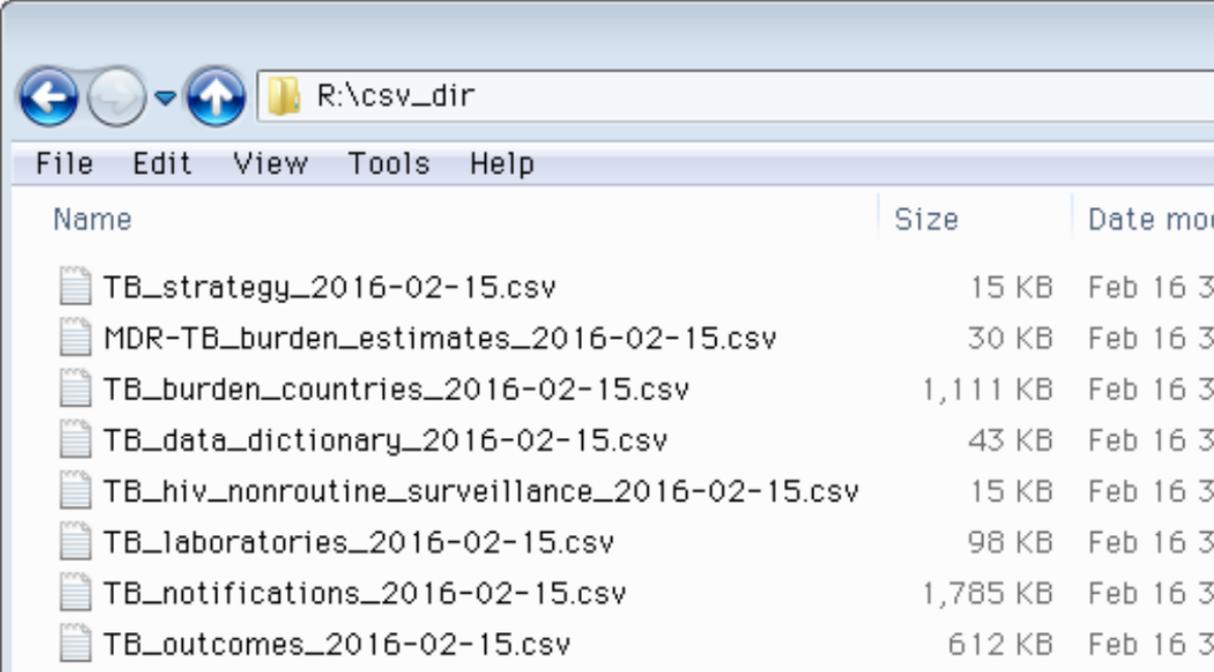
中身は普通のCSV



A screenshot of the AkelPad text editor displaying a CSV file named "TB_burden_countries_2016-02-15.csv". The window title is "AkelPad - [R:\csv_dir\TB_burden_countries_2016-02-15.csv]". The menu bar includes "File", "Edit", "View", "Options", "Window", and "Help". The main text area contains a CSV file with 17 columns and 17 rows of data for Afghanistan. The columns are: "country", "iso2", "iso3", "iso_numeric", "g_whoregion", "year", "e_pop_num", and seven unlabeled columns. The data shows a steady increase in population and a decrease in TB burden over the period.

country	iso2	iso3	iso_numeric	g_whoregion	year	e_pop_num										
Afghanistan	AF	AFG	004	EMR	1990	12067570	307	160	502	37000	190					
Afghanistan	AF	AFG	004	EMR	1991	12789374	344	180	562	44000	230					
Afghanistan	AF	AFG	004	EMR	1992	13745630	373	191	613	51000	260					
Afghanistan	AF	AFG	004	EMR	1993	14824371	392	194	657	58000	290					
Afghanistan	AF	AFG	004	EMR	1994	15869967	409	198	694	65000	310					
Afghanistan	AF	AFG	004	EMR	1995	16772522	425	201	731	71000	340					
Afghanistan	AF	AFG	004	EMR	1996	17481800	436	202	758	76000	350					
Afghanistan	AF	AFG	004	EMR	1997	18034130	444	204	776	80000	370					
Afghanistan	AF	AFG	004	EMR	1998	18511480	446	204	781	83000	380					
Afghanistan	AF	AFG	004	EMR	1999	19038420	443	203	774	84000	390					
Afghanistan	AF	AFG	004	EMR	2000	19701940	431	203	743	85000	400					
Afghanistan	AF	AFG	004	EMR	2001	20531160	410	198	697	84000	410					
Afghanistan	AF	AFG	004	EMR	2002	21487079	393	196	658	84000	420					
Afghanistan	AF	AFG	004	EMR	2003	22507368	376	191	622	85000	430					
Afghanistan	AF	AFG	004	EMR	2004	23499850	359	186	586	84000	440					
Afghanistan	AF	AFG	004	EMR	2005	24399948	346	181	563	84000	440					
Afghanistan	AF	AFG	004	EMR	2006	25183615	338	177	548	85000	450					

スキーマと見なす場所に置く



A screenshot of a Windows Explorer window showing a directory named "R:\csv_dir". The window title bar includes navigation buttons and the address bar. The menu bar contains "File", "Edit", "View", "Tools", and "Help". The main area displays a list of CSV files with columns for "Name", "Size", and "Date modified".

Name	Size	Date modified
TB_strategy_2016-02-15.csv	15 KB	Feb 16 3
MDR-TB_burden_estimates_2016-02-15.csv	30 KB	Feb 16 3
TB_burden_countries_2016-02-15.csv	1,111 KB	Feb 16 3
TB_data_dictionary_2016-02-15.csv	43 KB	Feb 16 3
TB_hiv_nonroutine_surveillance_2016-02-15.csv	15 KB	Feb 16 3
TB_laboratories_2016-02-15.csv	98 KB	Feb 16 3
TB_notifications_2016-02-15.csv	1,785 KB	Feb 16 3
TB_outcomes_2016-02-15.csv	612 KB	Feb 16 3

クエリで Import Foreign Schema

cygwin

```
# create schema csv_who;
CREATE SCHEMA
#
# import foreign schema ogr_all
#   from server csv_dir
#   into csv_who;
NOTICE: Check schema 0 ogr_all
NOTICE: Number of tables to be created 8
IMPORT FOREIGN SCHEMA
#
```

自動的に8つの外部テーブルができてる

```
cygwin
```

```
# %det csv_who.*
```

```
List of foreign tables
```

```
+-----+-----+-----+
| Schema | Table | Server |
+-----+-----+-----+
| csv_who | mdr_tb_burden_estimates_2016_02_15 | csv_dir |
| csv_who | tb_burden_countries_2016_02_15 | csv_dir |
| csv_who | tb_data_dictionary_2016_02_15 | csv_dir |
| csv_who | tb_hiv_nonroutine_surveillance_2016_02_15 | csv_dir |
| csv_who | tb_laboratories_2016_02_15 | csv_dir |
| csv_who | tb_notifications_2016_02_15 | csv_dir |
| csv_who | tb_outcomes_2016_02_15 | csv_dir |
| csv_who | tb_strategy_2016_02_15 | csv_dir |
+-----+-----+-----+
(8 rows)
```

列名も CSV ヘッダから自動設定されてる

```
cygwin
# select * from csv_who.tb_strategy_2016_02_15;
+-[ RECORD 1 ]-----+
| fid          | 1
| country     | Afghanistan
| iso2        | AF
| iso3        | AFG
| iso_numeric  | 004
| g_whoregion | EMR
| year        | 2014
| xpert_in_guide_tbhiv | 0
| xpert_in_guide_mdr  | 1
| xpert_in_guide_tb   | 0
| xpert_in_guide_children | 0
| xpert_in_guide_eptb | 0
| eol_care           | 0
| caseb_err_nat      | 0
| hcw_tb_infected    |
| hcw_tot            | 30193
| pub_new_dx         | 366
| priv_new_dx        | 2826
```

日本語の实在データ（国勢調査 CSV）では

<https://www.e-stat.go.jp/SG1/estat>

[/GL08020103.do?_toGL08020103_&tclassID=000001034991&cycleCode=0](https://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&tclassID=000001034991&cycleCode=0)

統計表一覧 政府統計の | x

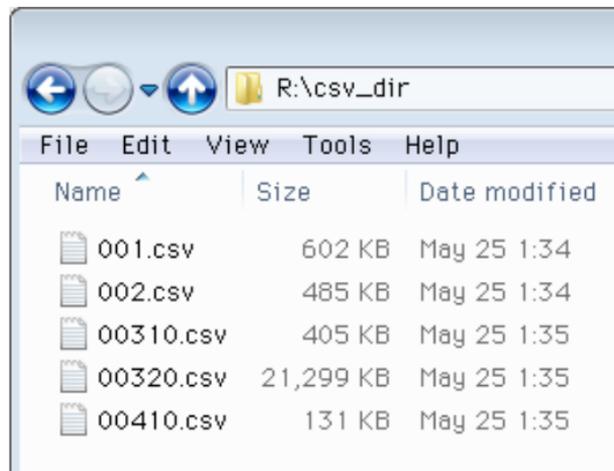
← → ↻ https://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&tclassID=00000103 ☆

平成22年国勢調査 > 人口等基本集計（男女・年齢・配偶関係、世帯の構成、住居の状態など） > 全国結果

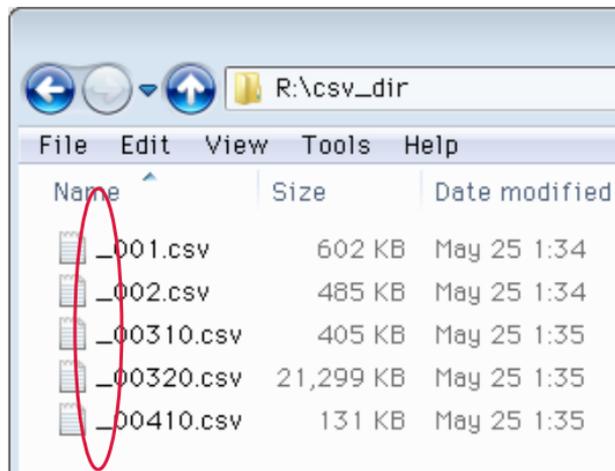
2011年10月26日 公

表番号	統計表		
総人口・総世帯数			
1	人口、人口増減、面積及び人口密度—全国※、全国市部※、全国郡部※、都道府県※、市部※、郡部※、市町村※・旧市町村	CSV	DB
2	男女別人口及び世帯の種類(2区分)別世帯数—全国※、全国市部※、全国郡部※、都道府県※、市部※、郡部※、市町村※・旧市町村	CSV	DB
男女・年齢・配偶関係			
3-1	年齢(各歳)、男女別人口、年齢別割合、平均年齢及び年齢中位数(総数及び日本人)—全国※、全国市部※、全国郡部※、都道府県※、市部※、郡部※、市町村※・旧市町村	CSV	DB
3-2	年齢(各歳)、男女別人口、年齢別割合、平均年齢及び年齢中位数(総数及び日本人)—全国※、全国市部※、全国郡部※、都道府県※、市部※、郡部※、市町村※・旧市町村	CSV	DB
4-1	年齢(各歳)、出生の月(4区分)、男女別人口(総数及び日本人)—全国※、全国市部、全国郡部	CSV	DB
4-2	年齢(各歳)、出生の月(4区分)、男女別人口(総数及び日本人)—全国※、全国市部、全国郡部、都道府県、20大都市	CSV	DB

さっきと同じ場所にダウンロード



大丈夫! ファイル名を変えれば



大丈夫! ファイル名を変えれば

File Explorer window showing the directory R:\csv_dir with the following files:

Name	Size
_001.csv	60
_002.csv	48
_00310.csv	40
_00320.csv	21,29
_00410.csv	13

Cygwin terminal output:

```
# import foreign schema ogr_all
  from server csv_dir into csv_estat;
NOTICE: Check schema 0 ogr_all
NOTICE: Number of tables to be created 5
IMPORT FOREIGN SCHEMA
#
# %det+ csv_estat.*
                                List of foreign tables
+-----+-----+-----+-----+-----+
| Schema | Table | Server | FDW Options | Description |
+-----+-----+-----+-----+-----+
| csv_estat | _001 | csv_dir | (layer '001') | |
| csv_estat | _002 | csv_dir | (layer '002') | |
| csv_estat | _00310 | csv_dir | (layer '00310') | |
| csv_estat | _00320 | csv_dir | (layer '00320') | |
| csv_estat | _00410 | csv_dir | (layer '00410') | |
+-----+-----+-----+-----+-----+
(5 rows)
```

ヘッダ行がなければ、列名は field1, field2, ... と自動で付く

```
cygwin
# %d csv_estat.*
          Foreign table "csv_estat._001"
+-----+-----+-----+-----+
| Column |      Type      | Modifiers | FDW Options |
+-----+-----+-----+-----+
| fid    | integer        |           |             |
| field_1| character varying|         |             |
| field_2| character varying|         |             |
| field_3| character varying|         |             |
| field_4| character varying|         |             |
| field_5| character varying|         |             |
| field_6| character varying|         |             |
| field_7| character varying|         |             |
| field_8| character varying|         |             |
| field_9| character varying|         |             |
| field_10| character varying|         |             |
| field_11| character varying|         |             |
```

日本語も問題なく、外部テーブル化できた

cygwin

```
# select * from csv_estat._001 offset 10;
```

fid	field_1	field_2	field_3	field_4	field_5	field_6	field_7
11	11		00000	a	2010	2000	全国
12	12		00001	b	2010	2000	全国 市部
13	13		00002	b	2010	2000	全国 郡部
14	14		01000	a	2010	2000	北海道
15	15		01001	b	2010	2000	北海道 市部
16	16		01002	b	2010	2000	北海道 郡部
17	17		01100	1	2010	2000	札幌市
18	18		01101	0	2010	2000	札幌市 中央区
19	19		01102	0	2010	2000	札幌市 北区
20	20		01103	0	2010	2000	札幌市 東区
21	21		01104	0	2010	2000	札幌市 白石区
22	22		01105	0	2010	2000	札幌市 豊平区
23	23		01106	0	2010	2000	札幌市 南区
24	24		01107	0	2010	2000	札幌市 西区
25	25		01108	0	2010	2000	札幌市 厚別区
26	26		01109	0	2010	2000	札幌市 手稲区
27	27		01110	0	2010	2000	札幌市 清田区
28	28		01202	2	2010		函館市

日本語も問題なく、外部テーブル化できた

d_7	field_8	field_9	field_10	field_11	field_12	field_13
	128057352	127767994	289358	0.2264714278	377950.10	343.4
	118158631	115503691	652940	0.5852979523	216209.27	537.5
	11900721	12264303	-363582	-2.9645549364	161654.94	75.9
	5506419	5627737	-121318	-2.1557155212	83456.87	70.2
	4449360	4503624	-54264	-1.2048963235	18540.30	241.3
	1057059	1124113	-67054	-5.9650586729	64858.83	17.6
	1913545	1880863	32682	1.7376066199	1121.12	1706.8
	220189	202801	17388	8.5739222193	46.42	4743.4
	278781	272877	5904	2.1636121769	63.48	4391.6
	255873	253996	1877	0.7389880156	57.13	4478.8
	204259	201307	2952	1.4664169651	34.58	5906.9
	212118	209428	2690	1.2844509808	46.35	4576.4
	146341	153021	-6680	-4.3654138974	657.23	222.7
	211229	207329	3900	1.8810682538	74.93	2819.0
	128492	129720	-1228	-0.9466543324	24.38	5270.4
	139644	137601	2043	1.4847275819	56.92	2453.3
	116619	112783	3836	3.4012218153	59.70	1953.4
	279127	294264	-15137	-5.1440203355	677.93	411.7

嵌まらないための注意点、ほか

- ファイル名の先頭はアルファベットか、アンダースコアにする
- 多言語データは UTF-8 で保存し直してから!
- 先頭行が「ヘッダ」と見なされない場合、列名は field1, field2, ...
- 左端に、行番号 ID の列 fid が自動で付く
- 列の型は、行番号 ID を除き全て文字型。
→ 数値セルに適宜データ型を指定する、等の後処理が必要

とは言え、フォルダ内 CSV を一括自動で外部テーブルにできるのは、間違いなく便利

Excelファイル(xls, xlsx)の場合

最初に、まとめ

	xls (～ Excel2003)	xlsx (Excel2007 ～)
ASCII だけの列	○	○
日本語を含む列	○	×
1 シートのファイル	○	○
複数シートのファイル	×	○

【注】 CSV と違って「フォルダ内一括処理」はできない

実際：CSVと同じく日本の国勢調査データ

http://www.e-stat.go.jp/SG1/estat

/GL08020103.do?_toGL08020103_&tclassID=000001005118&cycleCode=0

統計表一覧 政府統計の総合窓口

www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&tclassID=000001005118&cycleCode=0

平成17年国勢調査 > 男女・年齢・配偶関係、世帯の構成、住居の状態など（第1次基本集計） > 全国結果 > 報告書掲載表

2006年10月

表番号	統計表	
総人口・総世帯数		
1	人口、人口増減（平成12年～17年）、面積及び人口密度－全国※、市部※、郡部※、都道府県、15大都市	Excel
2	男女別人口及び世帯の種類（2区分）別世帯数－全国※、市部※、郡部※、都道府県、市部、郡部	Excel
男女・年齢・配偶関係		
3	年齢（各歳）、男女別人口、年齢別割合、平均年齢及び年齢中位数（総数及び日本人）－全国※、市部※、郡部※、都道府県、15大都市	
	総数	Excel
	日本人	Excel
	人口集中地区：総数	Excel
	人口集中地区：日本人	Excel

Excel 「データ」というか「統計表版下」

a001.xls [互換モード] - Microsoft Excel

ホーム 挿入 ページ レイアウト 数式 データ 校閲 表示 開発 Acrobat

O18 127767994

				人 口		平成12年～17年の人口増減		
地 域				平成17年	平成12年 (組替)	実 数	率 (%)	
Area				2005	2000	Number	Rate	
				(a)	(readjusted)			
全	国			Japan	127,767,994	126,925,843	842,151	
市	部			All shi	110,264,324	109,125,516	1,138,808	
郡	部			All gun	17,503,670	17,800,327	-296,657	-
01	北	海	道	Hokkaido	5,627,737	5,683,062	-55,325	-
02	青	森	県	Aomori-ken	1,436,657	1,475,728	-39,071	-
03	岩	手	県	Iwate-ken	1,385,041	1,416,180	-31,139	-
04	宮	城	県	Miyagi-ken	2,360,218	2,365,290	-5,072	-

第1表 人口, 人口増減(平成12年～17年), 面積及び人口密
- 全国*, 市部*, 郡部*, 都道府県, 15大都市

Table 1. Population, Population Change (2000-2005), Area and Population Densi
- Japan*, All Shi*, All Gun*, Prefectures and 15 Major Cities

数値のセルに、直接「注マーク」が入ってる

12年 組替)	実数	率(%)	Area	Population density
2000 adjusted)	Number	Rate	(b) 1)	(per km ³) (a) / (b)
126,925,843	842,151	0.7	2) 377,914.78	3)abc 342.7
109,125,516	1,138,808	1.0	4) 181,792.37	3)a 606.9
17,800,327	-296,657	-1.7	4) 195,025.66	3)bc 92.1
5,683,062	-55,325	-1.0	83,455.73	3)ab 71.8
1,475,728	-39,071	-2.6	4) 9,606.88	149.5
1,416,180	-31,139	-2.2	15,278.71	90.7
2,365,320	-5,102	-0.2	4) 7,285.60	324.0
1,189,279	-43,778	-3.7	4) 11,612.22	98.6
1,244,147	-27,966	-2.2	4) 9,323.39	130.4
2,126,935	-35,616	-1.7	13,782.75	151.7
2,985,676	-10,509	-0.4	6,095.68	488.1
2,004,817	11,814	0.6	6,408.28	314.7
2,024,852	-717	-0.0	6,363.16	318.1
6,938,006	116,237	1.7	4) 3,797.30	1,857.7
5,926,285	130,177	2.2	4) 5,156.68	1,174.5

1 ファイルごとに CREATE SERVER

cygwin

```
# create server xls_test
  foreign data wrapper ogr_fdw
  options (datasource 'r:/xls(x)/a001.xls', format 'xls');
CREATE SERVER
#
# create schema xls;
CREATE SCHEMA
```

そして Import Foreign Schema

cygwin

```
# create server xls_test
  foreign data wrapper ogr_fdw
  options (datasource 'r:/xls(x)/a001.xls', format 'xls');
CREATE SERVER
#
# create schema xls;
CREATE SCHEMA
#
# import foreign schema ogr_all
#   from server xls_test into xls;
NOTICE: Check schema 0 ogr_all
NOTICE: Number of tables to be created 1
IMPORT FOREIGN SCHEMA
#
```

テーブル名・列名は自動設定される (CSVと同様)

cygwin				cygwin		
# %det xls.*				# %d xls.a001		
List of foreign tables				Foreign table "xls.a001"		
+-----+-----+-----+				+-----+-----+		
Schema	Table	Server	Column	Type	Modifiers	
+-----+-----+-----+				+-----+-----+		
xls	a001	xls_test	fid	integer		
+-----+-----+-----+				+-----+-----+		
(1 row)				field1	character varying	
				field2	character varying	
				field3	character varying	
				field4	character varying	
				field5	character varying	
				field6	character varying	
				field7	character varying	
				field8	character varying	
				field9	character varying	

外部テーブルへのクエリで、Excelデータが表示される。日本語も正常

```
cygwin
# select fid, field10, field12, field15, field16, field17, field18, field19, field20
# from xls.a001 offset 17 limit 50;
```

fid	field10	field12	field15	field16	field17	field18	field19	field20
18			127767994	126925843	842151	0.7	377914.78	342.7
19	市部		110264324	109125516	1138808	1	181792.37	606.9
20	郡部		17503670	17800327	-296657	-1.7	195025.66	92.0999
21	01	北海道	5627737	5683062	-55325	-1	83455.73	71.8
22	02	青森県	1436657	1475728	-39071	-2.6	9606.879999999999	149.5
23	03	岩手県	1385041	1416180	-31139	-2.2	15278.71	90.7
24	04	宮城県	2360218	2365320	-5102	-0.2	7285.6	324
25	05	秋田県	1145501	1189279	-43778	-3.7	11612.22	98.5999
26	06	山形県	1216181	1244147	-27966	-2.2	9323.389999999999	130.4
27	07	福島県	2091319	2126935	-35616	-1.7	13782.75	151.7
28	08	茨城県	2975167	2985676	-10509	-0.4	6095.68	488.1
29	09	栃木県	2016631	2004817	11814	0.6	6408.28	314.7
30	10	群馬県	2024135	2024852	-717	0	6363.16	318.1
31	11	埼玉県	7054243	6938006	116237	1.7	3797.3	1857.7
32	12	千葉県	6058462	5926285	130177	2.2	5156.68	1174.5

(再掲) Excelデータインポートまとめ

	xls (～ Excel2003)	xlsx (Excel2007～)
ASCII だけの列	○	○
日本語を含む列	○	×
1 シートのファイル	○	○
複数シートのファイル	×	○

このバツが逆なら、xls に一本化で済むのに…

(今後)

- OGR_FDOW は、MySQL や SQLite にも接続できる
- 「PostGIS のエクストラ」 的な位置付けにとどまるのは、もったいない
- 他にも便利な使い方を考えて、情報発信できればと思います! (… ブログも再開して)