

# 日本の気象観測データの (オープンなようでオープンでない) 話

27<sup>th</sup> June 2020, Tokyo.R #86  
Yuta Kanzawa @yutakanzawa



Data Science Senior Analyst at Janssen Pharmaceutical K.K., Tokyo  
A Family Company of Johnson & Johnson



# I am...

- 神沢雄大 Yuta Kanzawa (twitter: [@yutakanzawa](https://twitter.com/yutakanzawa))
- Data scientist at Janssen Japan, Tokyo
  - A pharmaceutical company of J&J
- Opera & wine lover
  - Wagner
  - Bourgogne
- 7 languages
  - Human: Japanese, English, German
  - Computer: R, Python, SAS, SQL



# アジェンダ

- 今日話すこと
  - オープンデータ
  - 気象庁の気象観測データのダウンロードサービス
  - 国土数値情報API (悲報)
- 今日話さないこと
  - 気象観測データ
  - 国土数値情報
  - Python


# 前回までの話： 国土数値情報APIのススメ

Previously on Tokyo.R #84

# (再掲) 始まり : 日本の1キロメッシュの年平均気温図を作りたい。



The screenshot shows the R-bloggers website interface. At the top is the logo for R-bloggers, which includes a stylized 'R' with an orange arc above it and the text '-bloggers'. Below the logo is the tagline 'R news and tutorials contributed by hundreds of R bloggers'. A navigation bar contains links for Home, About, RSS, add your blog!, Learn R, R jobs, and Contact us. The main content area features a 'WELCOME!' section with a 'Follow @rbloggers' button and a subscriber count of 81.9k. Below this is a paragraph about the site's content. To the right is a search bar for R-bloggers. The featured article is 'Temperature changes in Germany visualized in R' by Yongfu, Liao, dated November 5, 2019. It has 263 likes and share buttons for Facebook, Twitter, and LinkedIn. A text box contains a notice about the article's original publication on Jan Knappe's blog. Below the article is a 'Share' button for Facebook and a 'Tweet' button for Twitter. A 'MOST VISITED ARTICLES OF THE WEEK' list is also visible.




The screenshot shows the ZEIT ONLINE website. At the top is the logo for ZEIT ONLINE and a search bar. Below the logo is a navigation bar with categories like Politik, Gesellschaft, Wirtschaft, Kultur, Wissen, Digital, Campus, Arbeit, Entdecken, Sport, ZEITmagazin, Podcasts, and mehr. There are also buttons for 'SCHWERPUNKTE' with sub-buttons for DONALD TRUMP, KLIMAGIPFEL, and UKRAINE-KONFLIKT. The main content area features a section titled 'Klimawandel' with the headline 'Viel zu warm hier'. Below the headline is a large grid of small, colorful icons representing temperature changes. At the bottom of the article is a paragraph in German: 'Durch den Klimawandel hat sich Deutschland deutlich erwärmt. Wir zeigen, wie sich die Temperatur in Ihrer Gemeinde in den letzten 137 Jahren entwickelt hat.' To the right of this paragraph is a byline: 'Von Paul Blicke, Elena Erdmann, Flavio Dortana u. a.' and a link to '145 Kommentare'.

\*1 <https://www.r-bloggers.com/temperature-changes-in-germany-visualized-in-r/>

\*2 <https://www.zeit.de/wissen/umwelt/2019-12/klimawandel-globale-erwaermung-warming-stripes-wohnort>

# (再掲) Next steps

- 国土数値情報API
  - ダウンロード機能をライブラリ化してみたい (RやPython)。
  - To 中の人：正式化と対応データ拡充を是非お願いします！
- 日本の年平均気温図
  - 気象庁サイトからの気象データダウンロード：大変そう 
  - 意図的にAPIを用意していない。
  - 継時的な観測所のリストがない。
    - サイトのHTMLからマイニングできそう\*1。

\*1 [http://washitake.com/weather/amedas/obs\\_stations.md](http://washitake.com/weather/amedas/obs_stations.md)

# 気象データダウンロードサービス

Not so open data

# その前に、オープンデータとは？

- 日本政府の定義：『オープンデータ基本指針』\*1
  1. 営利目的、非営利目的を問わず二次利用可能なルールが適用されたもの
  2. 機械判読に適したもの
  3. 無償で利用できるもの
- 同指針のオープンデータに関する基本的ルール：公開環境\*1
  - 「各府省庁は、ウェブサイトで容易に検索・利用できる形でデータを公開する。特にニーズが高いと想定されるデータについては、利用者の利便性に加え、システムの負荷及び効率性の観点から、一括ダウンロードを可能とする仕組みの導入や、APIを通じた提供を推進する。」

\*1 [https://cio.go.jp/sites/default/files/uploads/documents/data\\_shishin.pdf](https://cio.go.jp/sites/default/files/uploads/documents/data_shishin.pdf)



# 参考：日本政府の公式オープンデータカタログサイト\*1

日本語 | English

The screenshot shows the homepage of the official Japanese Open Data Catalog (DATA GO.JP). At the top, there is a dark blue header with the logo 'DATA GO.JP データカタログサイト'. Below the header is a navigation menu with items: お知らせ, データ, データベースサイト一覧, 公共データ活用事例, コミュニケーション, and 開発者向け情報. A search bar is located below the menu with the placeholder text 'データセットを検索...'. Below the search bar is a link to the '意見受付コーナー' (Opinion Reception Corner) with the text 'にて、オープンデータの掲載に関する御要望等を受け付けております (匿名可)'. The main content area is divided into several sections: 'データ' (Data) with icons for HTML, PDF, XLSX, CSV, XML, and JPG; 'オープンデータの取組 (リンク集)' (Open Data Initiatives (Link Collection)) with a central 'OPENDATA LINKS' icon and links to policies, public data usage examples, and database site lists; 'コミュニケーション' (Communication) with speech bubble icons and links to the opinion reception corner and public response corners; '一億総活躍社会の実現' (Achievement of a Society where 100 million people are fully engaged) with links to childcare support, social security, and economic growth; and '地方公共団体データベースサイト一覧' (List of Local Government Database Sites) with a bar chart icon.

\*1 <https://www.data.go.jp/>

# 気象庁「過去の気象データ・ダウンロード」\*1

国土交通省  
気象庁  
Japan Meteorological Agency

ENGLISH  
Google カスタム検索 検索

ホーム 防災情報 各種データ・資料 知識・解説 気象庁について 案内・申請

ホーム > 各種データ・資料 > 過去の気象データ・ダウンロード

データ検索 過去の気象データ 地域平均気象 取り組んでみませんか? 気候リスク管理

過去の気象データ・ダウンロード 重要なお知らせ このページの使い方 よくある質問 GSVファイルの形式

検索条件 選択済みのデータ量 0% 100% (上限)

地点を選ぶ 項目を選ぶ 期間を選ぶ 表示オプションを選ぶ

すべての選択済みの地点をクリア

一回のリクエストで表示・ダウンロードできるデータ量には上限があります (右グラフ参照)。また、このページへのアクセスが集中したり、リクエストのデータ量が多い場合には、表示・ダウンロードまで時間がかかる場合があります。

まず、都道府県を選んでください

宗谷 留萌 上川 網走・北見・紋別 空知 石狩 日高 十勝 根室 釧路 檜山 胆振 渡島 青森 秋田 岩手 山形 宮城 福島 石川 富山 新潟 福井 滋賀 岐阜 長野 群馬 栃木 茨城 山梨 埼玉 東京 千葉 大阪 奈良 三重 愛知 静岡 神奈川 和歌山 三重 山口 島根 鳥取 兵庫 京都 滋賀 岐阜 長野 群馬 栃木 茨城 山梨 埼玉 東京 千葉 長崎 佐賀 福岡 熊本 大分 愛媛 香川 高知 徳島 鹿児島 宮崎 沖縄 南極

画面に表示

CSVファイルをダウンロード

選択地点・項目をクリア  
選択された地点 観測項目  
← 地点を選択してください

選択された項目  
← 項目を選択してください

選択された期間(日本標準時)  
2020年1月1日から  
2020年1月1日までの日別値を表示

\*1 <https://www.data.jma.go.jp/gmd/risk/obsdl/index.php>

# お気づきだろうか？

地点を選ぶ

項目を選ぶ

期間を選ぶ

表

す

一回のリクエストで表示・ダウンロードできるデータ量には上限があります  
(右上棒グラフ参照)。また、このページへのアクセスが集中したり、リクエストのデータ量が多い場合には、表示・ダウンロードまで時間がかかる場合があります。

後志

檜山

# ダウンロードしてみた：東京気象台の過去30年間の日次平均気温

過去の気象データ・ダウンロード

 [重要なお知らせ](#)

 [このページの使い方](#)

 [よくある質問](#)

 [CSVファイルの形式](#)

検索条件

選択済みのデータ量 0%  100% (上限)

地点を選ぶ

項目を選ぶ

期間を選ぶ

表示オプションを選ぶ

画面に表示 ▶

CSVファイルをダウンロード ▶

選択地点・項目をクリア

選択された地点 観測項目

東京



削除

選択された項目

日平均気温

削除

期間

連続した期間で表示する

最近1年

最近1か月

1990 年 6 月 25 日から

2020 年 6 月 25 日までの日別値を表示

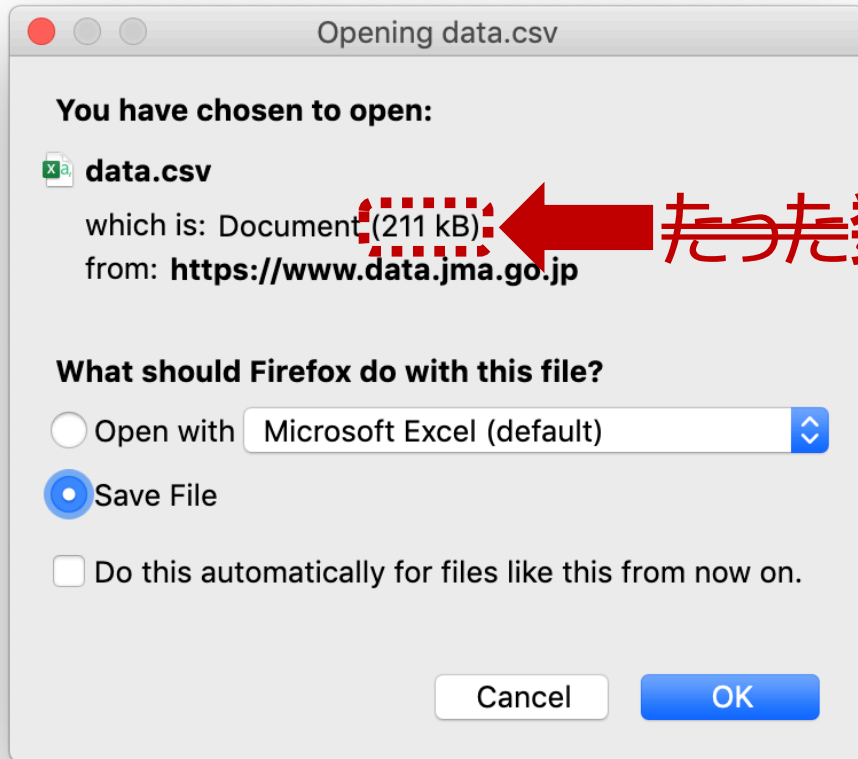
特定の期間を複数年分、表示する

1 月 1 日から 1 月 1 日の値を

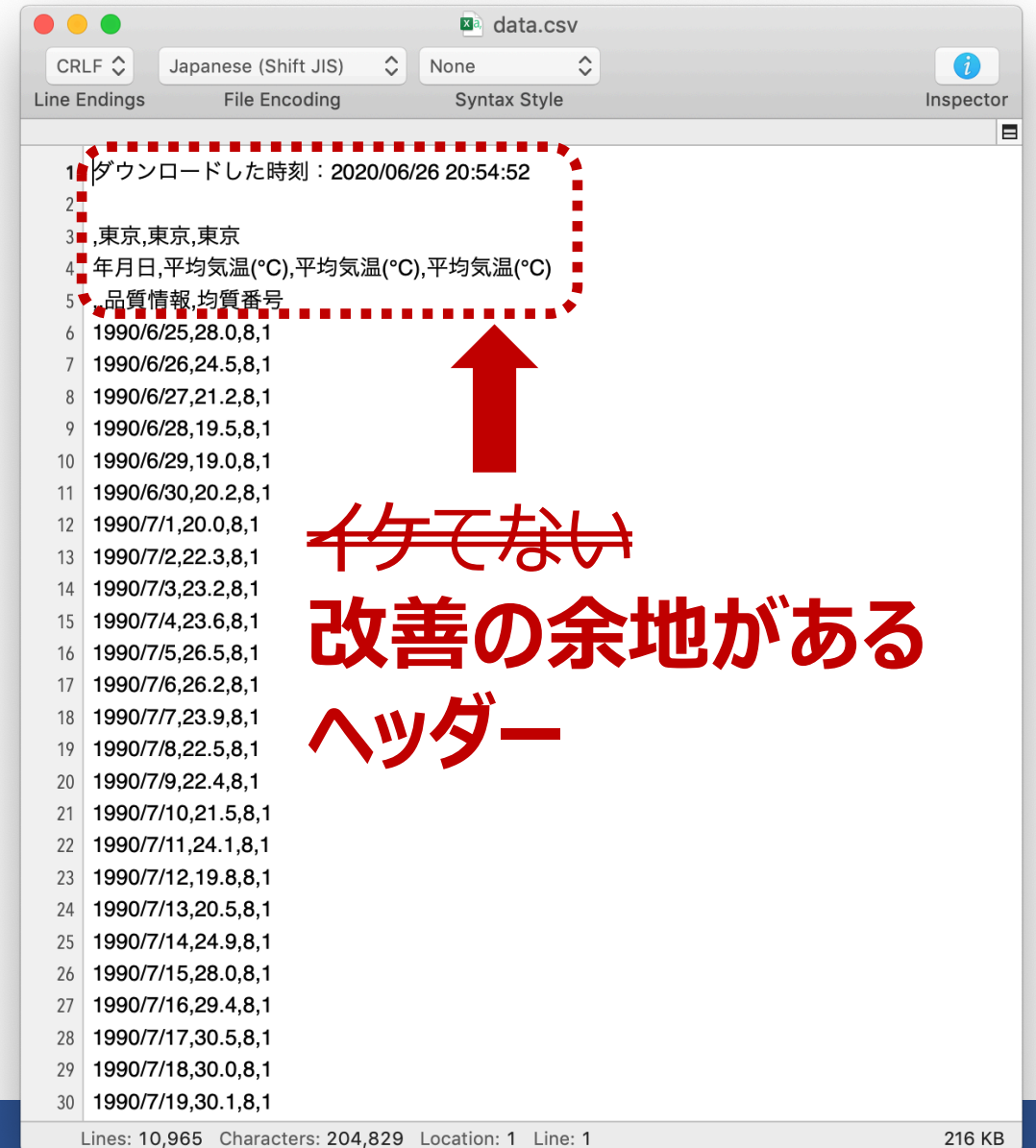
2020 年から 2020 年まで表示

 [期間選択の使い方](#)

# 結果：イマイチ...



たった数百KB



↑

~~イマイチ~~  
改善の余地がある  
ヘッダー

# なぜ上限がある???

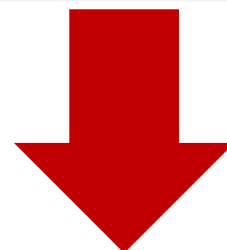
## • 「よくある質問」\*1へのご回答

取得できるデータ量に制限があって使いづらいのですが、まとめて取得する方法はありますか？

本ツールは、多くの方々が同時に利用しても円滑に動作するよう、一度に取得できるデータ量に制限を設けています。大容量のデータが必要な場合は、これらを販売している(一財)気象業務支援センターへお問い合わせください。



たった数百KB



~~天下り先へ利益誘導~~

気象ビッグデータのプロが有償対応！ 🤔

\*1 <https://www.data.jma.go.jp/gmd/risk/obsdl/top/faq.html>

# (再掲) その前に、オープンデータとは？

- 日本政府の定義：『オープンデータ基本指針』\*1
  1. 営利目的、非営利目的を問わず二次利用可能なルールが適用されたもの
  2. 機械判読に適したもの
  3. 無償で利用できるもの
- 『オープンデータに関する基本的ルール』：公開環境\*1
  - 「各府省庁は、ウェブサイトで容易に検索・利用できる形でデータを公開する。特にニーズが高いと想定されるデータについては、利用者の利便性に加え、システムの負荷及び効率性の観点から、一括ダウンロードを可能とする仕組みの導入や、APIを通じた提供を推進する。」

\*1 [https://cio.go.jp/sites/default/files/uploads/documents/data\\_shishin.pdf](https://cio.go.jp/sites/default/files/uploads/documents/data_shishin.pdf)

# 自分でオープンにしてみた：

## →気象庁のデータ検索ページ\*1からスクレイピング

ホーム > 各種データ・資料 > 過去の気象データ検索

過去の気象データ検索

各地の気温、降水量、風など | 高層の気温、風など

最新の気象データ | 過去の気象データダウンロード | 過去の地域平均気象データ

地点と年月日時を選択して、表示するデータの種類を選択してください。検索条件を全てクリア

地点の選択	年月日の選択	データの種類
<p>地点の選択をクリア</p> <p>都道府県</p> <p>都道府県・地方を選択</p> <p>地点</p> <p>地点を選択</p> <p>(未選択)</p>	<p>年月日の選択をクリア</p> <p>2020年 2000年 1980年</p> <p>2019年 1999年 1979年</p> <p>2018年 1998年 1978年</p> <p>2017年 1997年 1977年</p> <p>2016年 1996年 1976年</p> <p>2015年 1995年</p> <p>2014年 1994年</p> <p>2013年 1993年</p> <p>2012年 1992年</p> <p>2011年 1991年</p> <p>2010年 1990年</p> <p>2009年 1989年</p> <p>2008年 1988年</p> <p>2007年 1987年</p> <p>2006年 1986年</p> <p>2005年 1985年</p> <p>2004年 1984年</p> <p>2003年 1983年</p> <p>2002年 1982年</p> <p>2001年 1981年</p> <p>1月 1日 16日</p> <p>2月 2日 17日</p> <p>3月 3日 18日</p> <p>4月 4日 19日</p> <p>5月 5日 20日</p> <p>6月 6日 21日</p> <p>7月 7日 22日</p> <p>8月 8日 23日</p> <p>9月 9日 24日</p> <p>10月 10日 25日</p> <p>11月 11日 26日</p> <p>12月 12日 27日</p> <p>13日 28日</p> <p>14日 29日</p> <p>15日 30日</p> <p>31日</p>	<p>年ごとの値を表示 (地点を指定してください)</p> <p>3か月ごとの値を表示 (地点、年を指定してください)</p> <p>観測開始からの月ごとの値を表示 (地点を指定してください)</p> <p>月ごとの値を表示 (地点、年を指定してください)</p> <p>旬ごとの値を表示 (地点、年を指定してください)</p> <p>半旬ごとの値を表示 (地点、年を指定してください)</p> <p>日ごとの値を表示 (地点、年月を指定してください)</p> <p>1時間ごとの値を表示 (地点、年月日を指定してください)</p> <p>10分ごとの値を表示 (地点、年月日を指定してください)</p> <p>地点ごとの観測史上1~10位の値 (地点を指定してください)</p> <p>歴代全国ランキング</p> <p>年・月ごとの平均値を表示 (地点を指定してください)</p> <p>3か月ごとの平均値を表示 (地点を指定してください)</p> <p>旬ごとの平均値を表示 (地点を指定してください)</p> <p>半旬ごとの平均値を表示 (地点を指定してください)</p> <p>日ごとの平均値を表示 (地点、月を指定してください)</p> <p>霜・雪・結氷の初終日と初冠雪日の平均値を表示 (気象台、測候所などのみのデータです)</p> <p>年平均値は1981~2010年の30年間の観測値の平均をもとに算出しています。</p> <p>要素別データの公開期間 (気象台、測候所などのみのデータです)</p> <p>メッシュ年平均図</p>

```
library(tidyverse)
library(rvest)
library(lubridate)

obs <- readRDS("r_data/obs_stations.rdf")

obs_list <- mapply(list, obs$type, obs$prec_no, obs$block_no, SIMPLIFY=F)

base_url <- "https://www.data.jma.go.jp/obd/stats/etrn/view/"
df <- data.frame()
for (e in obs_list){
  type <- e[[1]]
  prec_no <- e[[2]]
  block_no <- e[[3]]
  print(block_no)
  page_url <- paste0("annually_", type, ".php?prec_no=", prec_no,
                    "&block_no=", block_no,
                    "&year=&month=&day=&view=a2")
  block <- paste0(base_url, page_url) %>% read_html()
  tbl <- block %>% html_node("[id='tablefix1']") %>%
    html_table(header = FALSE, fill = TRUE)
```

\*1 <https://www.data.jma.go.jp/obd/stats/etrn/index.php>



# (悲報) 国土数値情報API

Goodbye

# 国土数値情報APIわず (2020年4月)

GISホームページ  
国土交通省国土政策局国土情報課

ENGLISH

GISホームページへ

-----お知らせ-----

システムメンテナンスのため、ご迷惑をおかけしますが下記の期間につきましては、事務局へのお問い合わせのメールアドレスを一時的に下記に変更いたします。

問い合わせ先変更予定期間：令和2年4月1日（水）～令和2年4月17日（金）  
\*予定は前後する場合がございます。  
国土情報提供サイト問合せ受付：[hqt-nsdijp@ml.mlit.go.jp](mailto:hqt-nsdijp@ml.mlit.go.jp)

Home

初めての方へ

よくある質問 (FAQ)

サイトマップ

利用約款

問合せ

## 国土数値情報 ダウンロードサービス

Web API

「国土数値情報」とは、国土形成計画、国土利用計画の策定等の国土政策の推進に資するために、地形、土地利用、公共施設などの国土に関する基礎的な情報をGISデータとして整備したものです。そのうち公開に差し支えないものについて、「地理空間情報活用推進基本法」等を踏まえて無償で提供しています。

Google カスタム検索



「国土数値情報」は、全国計画・ブロック計画である国土形成計画、国土利用計画等の策定等の国土政策の推進に資することを目的に整備したものです。ただし、「国土数値情報」は、概ね国土地理院の2万5000分の1地形図（許容誤差：10m超）をベースに作成しており、データの一部は住所情報から街区レベルです。空間精度上、建物の判読やナビゲーションなどには適していません。  
また、タイムラグとして実況とは数年～5年程度の遅れがあり、更新頻度が低いデータもあります。さらに、収集可能な原典情報のみから整備していることによるデータの欠落があります。利用に当たっては、データの整備年月や著作権者等の許諾条件（ライセンス）に留意して使用いただくとともに、最新の情報が必要な場合はそれぞれ別途、原典の資料で確認するようにしてください。

### ■重大な警告■

本サイトで提供されているデータ（データセットやリソースの説明、API利用等）は、ジャンプしてきたリンク先やフレーム外の記載によらず、「国土数値情報」の利用約款等の本ホームページの記載のみに依拠し、それ以外の記載は全て無効です。ご注意ください。

「国土数値情報」はGISで利用できる形式（テキスト形式、シェープファイル形式、XML(JPGIS1.0またはJPGIS2.1(GML))形式の3形式）で配信しています（ご利用にはGISソフトウェアが必要となります）。

「国土数値情報」は「[国土情報ウェブマッピングシステム](#)」においてウェブブラウザ上で閲覧可能です。

初めての方へ

国土数値情報 [パンフ](#)

[レット](#) ●新規

国土数値情報 [閲覧マ](#)

[形式 \(GML\)](#)

データ形式	JPGIS形式		旧 統一フォーマット形式	
	GML(JPGIS2.1)シェープファイル	XML(JPGIS1.0)	GML(JPGIS2.1)シェープファイル	テキスト
	1. 国土 (水・土地)			

基本情報

[局の概要・組織図 \(PDF\)](#)

あった

\*1 <http://nlftp.mlit.go.jp/ksj/index.html>

# 国土数値情報APIなう（2020年6月）

🏠 トップページへ

🗺️ 国土数値情報ダウンロード

🗺️ 位置参照情報ダウンロード

🗺️ 国土調査（土地分類調査・水調査）

🗺️ ウェブマッピング

GML(JPGIS2.1)シェープファイル

- 1. 国土（水・土地）
- 2. 政策区域
- 3. 地域
- 4. 交通
- 5. 各種統計

XML(JPGIS1.0)

GML(JPGIS2.1)シェープファイル

テキスト

国土数値情報ダウンロード



ENHANCED BY Google



もうない！

データ形式

JPGIS形式

旧 統一フォーマット形式

GML (JPGIS2.1)  
シェープファイル

XML (JPGIS1.0)

GML (JPGIS2.1)  
シェープファイル

テキスト

1. 国土（水・土地）

水域

海岸線（ライン）

海岸保全施設（ライン）（ポイント）

湖沼（ポリゴン）

流域メッシュ

ダム（ポイント）

河川（ライン）（ポイント）



\*1 <http://nlftp.mlit.go.jp/ksj/index.html>

# RIP...

- 「API」でサイト内検索をしても0件。
- kokudosuuchiライブラリーはエラーが返ってくる。

```
> library(kokudosuuchi)
```

このサービスは、「国土交通省 国土数値情報（カテゴリ名）」をもとに加工者が作成  
以下の国土数値情報ダウンロードサービスの利用約款をご確認の上ご利用ください：

```
http://nlftp.mlit.go.jp/ksj/other/yakkan.html
```

```
> getKSJSummary()
```

```
read_xml.raw(charToRaw(enc2utf8(x)), "UTF-8", ..., as_html = as_html, でエラー:  
EntityRef: expecting ';' [23]
```

```
> |
```

# まとめ

Long story short

# Next steps

- 気象庁の気象データダウンロードサービス
  - To 中の人：もっとオープンにしてください！
- 国土数値情報API
  - To 中の人：復活させてください！！！！
- 日本の年平均気温図
  - モデリング
    - まずは単純な重回帰

**Enjoy!**