



# サービス開発と R パッケージ開発の話

2020年12月5日

株式会社コネクトデータ

安部 晃生

[@kos59125](https://twitter.com/kos59125)



<https://connecto-data.com>  
[contact@connecto-data.com](mailto:contact@connecto-data.com)

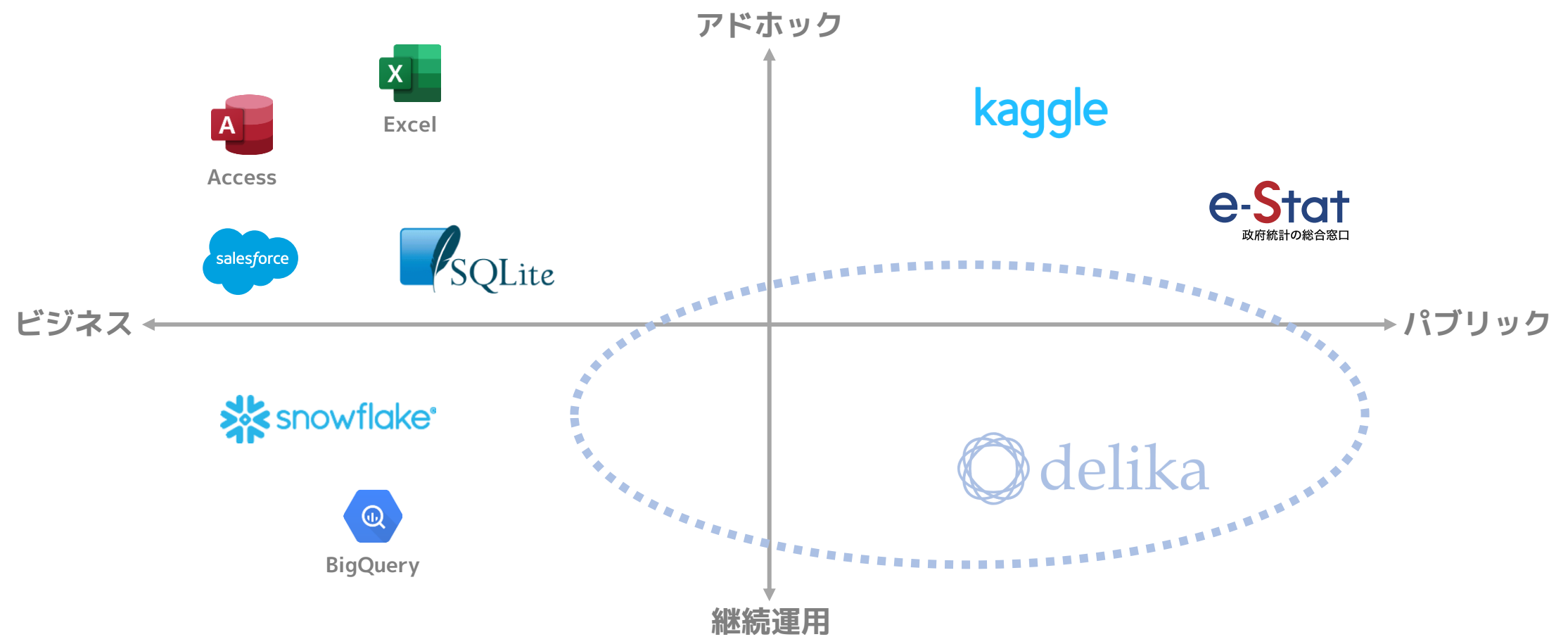
## I. 作ったサービスの話

## II. 作った R パッケージの話

- i. 配布方法の検討
- ii. 配布までの流れ

## III. まとめ

簡単にデータ共有、データ利用（DL/SQL）ができるオープンな DWH サービス。分析用のデータの保存やスクレイピングデータの公開に。



# スクレイピングと著作権の話

## 日本の著作権法はスクレイピングして作った学習用データやそのデータを利用した結果を勝手に公開しても良いと言っている。

### 第三十条の四

著作物は、次に掲げる場合その他の当該著作物に表現された思想又は感情を自ら享受し又は他人に享受させることを目的としない場合には、その必要と認められる限度において、いずれの方法によるかを問わず、利用することができる。ただし、当該著作物の種類及び用途並びに当該利用の態様に照らし著作権者の利益を不当に害することとなる場合は、この限りでない。

- 一 著作物の録音、録画その他の利用に係る技術の開発又は実用化のための試験の用に供する場合
- 二 情報解析（多数の著作物その他の大量の情報から、当該情報を構成する言語、音、影像その他の要素に係る情報を抽出し、比較、分類その他の解析を行うことをいう。第四十七条の五第一項第二号において同じ。）の用に供する場合
- 三 前二号に掲げる場合のほか、著作物の表現についての人の知覚による認識を伴うことなく当該著作物を電子計算機による情報処理の過程における利用その他の利用（プログラムの著作物にあつては、当該著作物の電子計算機における実行を除く。）に供する場合

### 第四十七条の四

電子計算機における利用（情報通信の技術を利用する方法による利用を含む。以下この条において同じ。）に供される著作物は、次に掲げる場合その他これらと同様に当該著作物の電子計算機における利用を円滑又は効率的に行うために当該電子計算機における利用に付随する利用に供することを目的とする場合には、その必要と認められる限度において、いずれの方法によるかを問わず、利用することができる。ただし、当該著作物の種類及び用途並びに当該利用の態様に照らし著作権者の利益を不当に害することとなる場合は、この限りでない。

- 一 電子計算機において、著作物を当該著作物の複製物を用いて利用する場合又は無線通信若しくは有線電気通信の送信がされる著作物を当該送信を受信して利用する場合において、これらの利用のための当該電子計算機による情報処理の過程において、当該情報処理を円滑又は効率的に行うために当該著作物を当該電子計算機の記録媒体に記録するとき。
- 二 自動公衆送信装置を他人の自動公衆送信の用に供することを業として行う者が、当該他人の自動公衆送信の遅滞若しくは障害を防止し、又は送信可能化された著作物の自動公衆送信を中継するための送信を効率的に行うために、これらの自動公衆送信のために送信可能化された著作物を記録媒体に記録する場合
- 三 情報通信の技術を利用する方法により情報を提供する場合において、当該提供を円滑又は効率的に行うための準備に必要な電子計算機による情報処理を行うことを目的として記録媒体への記録又は翻案を行うとき。
- 四 電子計算機における利用に供される著作物は、次に掲げる場合その他これらと同様に当該著作物の電子計算機における利用を行うことができる状態を維持し、又は当該状態に回復することを目的とする場合には、その必要と認められる限度において、いずれの方法によるかを問わず、利用することができる。ただし、当該著作物の種類及び用途並びに当該利用の態様に照らし著作権者の利益を不当に害することとなる場合は、この限りでない。
- 一 記録媒体を内蔵する機器の保守又は修理を行うために当該機器に内蔵する記録媒体（以下この号及び次号において「内蔵記録媒体」という。）に記録されている著作物を当該内蔵記録媒体以外の記録媒体に一時的に記録し、及び当該保守又は修理の後に、当該内蔵記録媒体に記録する場合
- 二 記録媒体を内蔵する機器をこれと同様の機能を有する機器と交換するためにその内蔵記録媒体に記録されている著作物を当該内蔵記録媒体以外の記録媒体に一時的に記録し、及び当該同様の機能を有する機器の内蔵記録媒体に記録する場合
- 三 自動公衆送信装置を他人の自動公衆送信の用に供することを業として行う者が、当該自動公衆送信装置により送信可能化された著作物の複製物が滅失し、又は毀損した場合の復旧の用に供するために当該著作物を記録媒体に記録するとき。

### 第四十七条の五

電子計算機を用いた情報処理により新たな知見又は情報を創出することによつて著作物の利用の促進に資する次の各号に掲げる行為を行う者（当該行為の一部を行う者を含み、当該行為を政令で定める基準に従つて行う者に限る。）は、公衆への提供又は提示（送信可能化を含む。以下この条において同じ。）が行われた著作物（以下この条及び次条第二項第二号において「公衆提供提示著作物」という。）（公表された著作物又は送信可能化された著作物に限る。）について、当該各号に掲げる行為の目的上必要と認められる限度において、当該行為に付随して、いずれの方法によるかを問わず、利用（当該公衆提供提示著作物のうちその利用に供される部分の占める割合、その利用に供される部分の量、その利用に供される際の表示の精度その他の要素に照らし軽微なものに限る。以下この条において「軽微利用」という。）を行うことができる。ただし、当該公衆提供提示著作物に係る公衆への提供又は提示が著作権を侵害するものであること（国外で行われた公衆への提供又は提示にあつては、国内で行われたとしたならば著作権の侵害となるべきものであること）を知りながら当該軽微利用を行う場合その他当該公衆提供提示著作物の種類及び用途並びに当該軽微利用の態様に照らし著作権者の利益を不当に害することとなる場合は、この限りでない。

- 一 電子計算機を用いて、検索により求める情報（以下この号において「検索情報」という。）が記録された著作物の題号又は著作者名、送信可能化された検索情報に係る送信元識別符号（自動公衆送信の送信元を識別するための文字、番号、記号その他の符号をいう。）その他の検索情報の特定又は所在に関する情報を検索し、及びその結果を提供すること。
- 二 電子計算機による情報解析を行い、及びその結果を提供すること。
- 三 前二号に掲げるもののほか、電子計算機による情報処理により、新たな知見又は情報を創出し、及びその結果を提供する行為であつて、国民生活の利便性の向上に寄与するものとして政令で定めるもの
- 四 前項各号に掲げる行為の準備を行う者（当該行為の準備のための情報の収集、整理及び提供を政令で定める基準に従つて行う者に限る。）は、公衆提供提示著作物について、同項の規定による軽微利用の準備のために必要と認められる限度において、複製若しくは公衆送信（自動公衆送信の場合にあつては、送信可能化を含む。以下この項及び次条第二項第二号において同じ。）を行い、又はその複製物による頒布を行うことができる。ただし、当該公衆提供提示著作物の種類及び用途並びに当該複製又は頒布の部数及び当該複製、公衆送信又は頒布の態様に照らし著作権者の利益を不当に害することとなる場合は、この限りでない。

### 施行令 第七条の四

第四十七条の五第一項（法第八十六条第一項及び第三項並びに第百二条第一項において準用する場合を含む。第三号において同じ。）の政令で定める基準は、次のとおりとする。

- 一 送信可能化された検索情報に係る送信元識別符号を検索し、及びその結果を提供する行為（□及び次項第一号において「送信元識別符号検索結果提供」という。）を行う場合にあつては、次に掲げる要件に適合すること。
- イ 送信可能化された著作物等に係る自動公衆送信について受信者を識別するための情報の入力を求めることその他の受信を制限するための手段が講じられている場合にあつては、当該自動公衆送信の受信について当該手段を講じた者の承諾を得たものに限つて利用を行うこと。
- ロ イに掲げるもののほか、送信元識別符号検索結果提供を適正に行うために必要な措置として文部科学省令で定める措置を講ずること。
- 二 第四十七条の五第二項（法第八十六条第一項及び第三項並びに第百二条第一項において準用する場合を含む。次項において同じ。）の規定の適用を受けて作成された著作物等の複製物を使用する場合にあつては、当該複製物に係る情報の漏えいの防止のために必要な措置を講ずること。
- 三 前二号に掲げるもののほか、法第四十七条の五第一項各号に掲げる行為に係る著作物等の利用を適正に行うために必要な措置として文部科学省令で定める措置を講ずること。
- 四 法第四十七条の五第二項の政令で定める基準は、次のとおりとする。
- 一 送信元識別符号検索結果提供を行う場合にあつては、当該送信元識別符号検索結果提供を前項第一号に掲げる要件に適合させるために必要な措置を講ずること。
- 二 法第四十七条の五第二項の規定の適用を受けて作成された著作物等の複製物に係る情報の漏えいの防止のために必要な措置を講ずること。

# スクレイピングと著作権の話

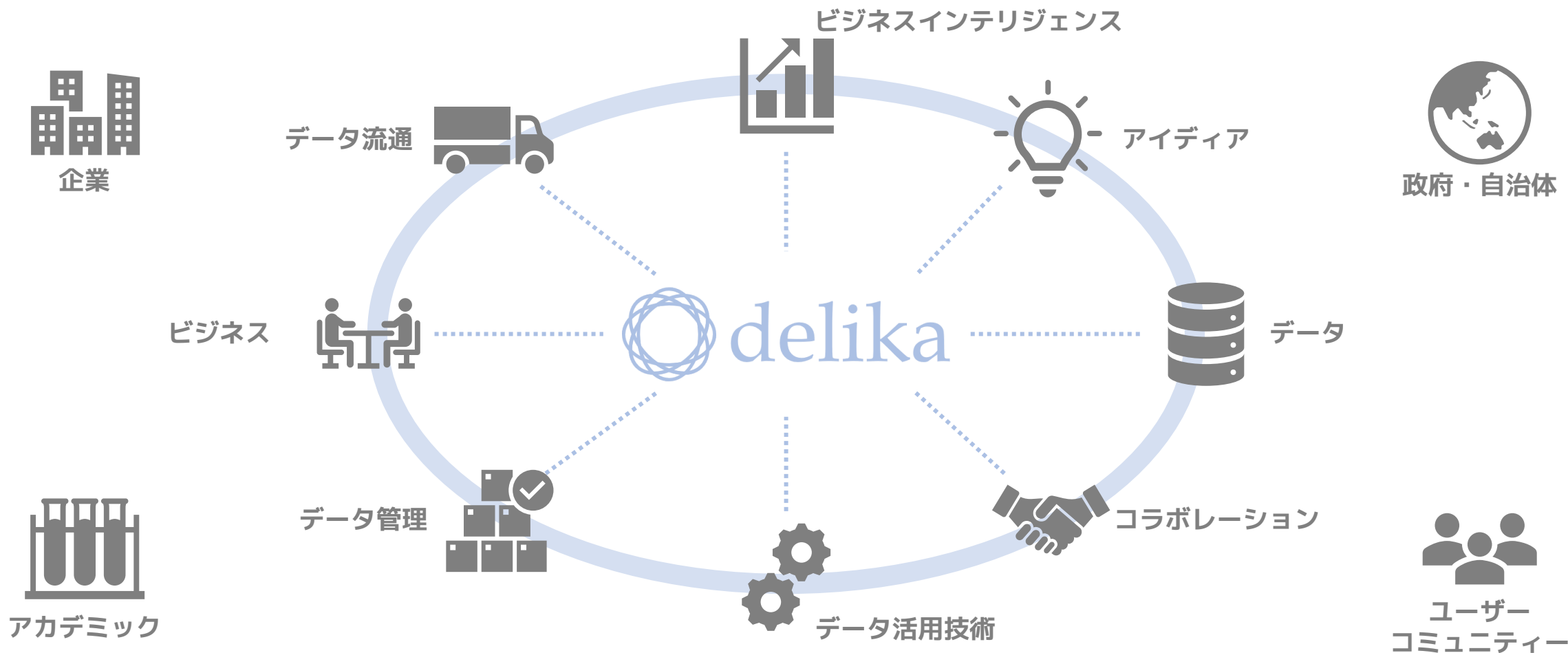
日本の著作権法はスクレイピングして作った学習用データやそのデータを利用した結果を勝手に公開しても良いと言っている。

著作権法（施行令）	大体こんなことが書いてある※
第三十条の四	著作物は、学習データとして使ってもいいよ。
第四十七条の四第一項	著作物は、前処理して保存してもいいよ。
第四十七条の五第一項	公開されている著作物は、分析してその結果を公開してもいいよ。
第四十七条の五第二項	公開されている著作物の分析結果を出す過程でできる中間データは、公開して配布してもいいよ。
施行令第七条の四第一項	公開されているといっても、ログインしないと見られなかったり robots.txt でスクレイピングが禁止されているような情報の利用は、許可が必要だよ。

※ 条文の解釈の部分的な抜粋をデータ分野に寄せたもの。正しく理解するためには参考文献を参照のこと。

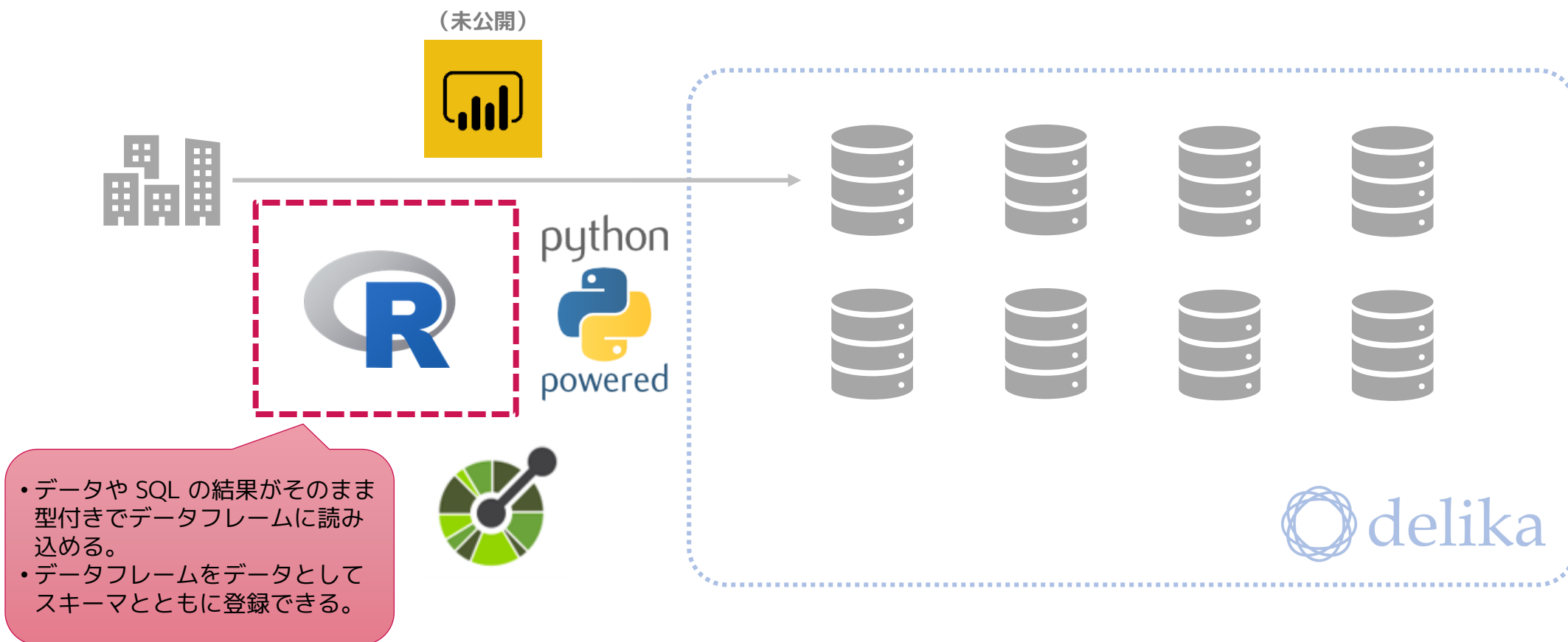
# 「delika」が目指しているもの

個人、法人を問わず、delika を中心としたデータ活用のコミュニティがデータ活用を発展させていくパブリックなデータインフラの提供。



# 今日の本題

delika は 自社ドキュメントサイト上で R クライアントを公開 している。  
なぜそのようにしたか、どう実現したか。



# パッケージの公開方法の選択

R パッケージ単体で開発しているのではなく、R パッケージは delika サービスの一部。サービス全体として構成管理することを重視した。

配布方法	運用が楽	公開が簡単	自由	認知獲得	導入容易性	サポート	統一感
公式 CRAN	○	×	×	○	○	△	×
<i>delika's choice</i> セルフホスト CRAN	△	△	○	×	△	○	○
GitHub 等	○	○	△	△	△	△	△
プライベート配布	△	△	○	×	×	○	○



# セルフホスト CRAN

セルフホスト CRAN は、特定のディレクトリー構造でウェブサイト上に配置するだけ。

```
/
/src/contrib
  package_1.0.tar.gz
  PACKAGES.rds
  PACKAGES.gz
  PACKAGES
  /PATH
    package_1.2.tar.gz
  /Archive/
    /package/
      package_0.9.tar.gz
/bin
/windows/contrib
  /3.3
  /3.4
  /PACKAGES
  /package_1.0.zip
/macosx/
/contrib
  /3.3
  /3.4
  /PACKAGES
  /package_1.0.tgz
/mavericks
/leopard
/el-capitan
```



バイナリ配布しないなら最低限これだけあれば OK

依存パッケージは公式 CRAN を参照しつつ独自パッケージを配布できる

```
# choose your favorite CRAN instead of https://cloud.r-project.org
repos <- c("https://cloud.r-project.org", "https://docs.delika.io/R")
# install the package and its dependencies.
install.packages("delika", repos = repos)
```

インストールスクリプト（「[delika R Client](#)」より）

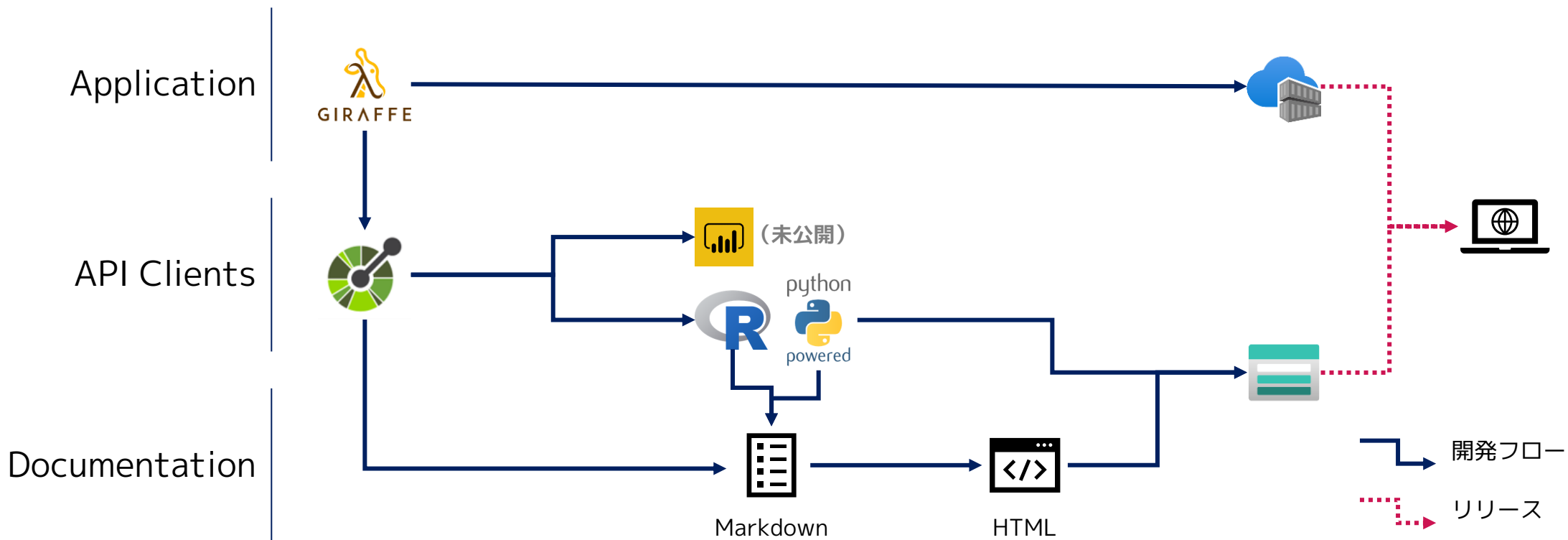
セルフホスト CRAN と公式 CRAN を両方指定することで、セルフホスト CRAN で提供していない依存パッケージが取得できる。

ディレクトリー構造（「[Environments: Repositories](#)」より）

# API クライアントの開発の流れ

delika の API クライアント開発はサービス本体の開発、ドキュメントサイトの開発等とともに Azure DevOps 上で行っている。

## delika の API 開発フロー (概要)



## まとめ

- R パッケージの配布方法には多様な選択肢があり、目的に応じて最適な方法を選択するべき。
- delika はサービスの構成管理のために一元的に管理できるようにセルフホストを選択した。
- delika 使ってね。

## 主要参考文献

- [「デジタル化・ネットワーク化の進展に対応した柔軟な権利制限規定に関する基本的な考え方」](#)（文化庁）

