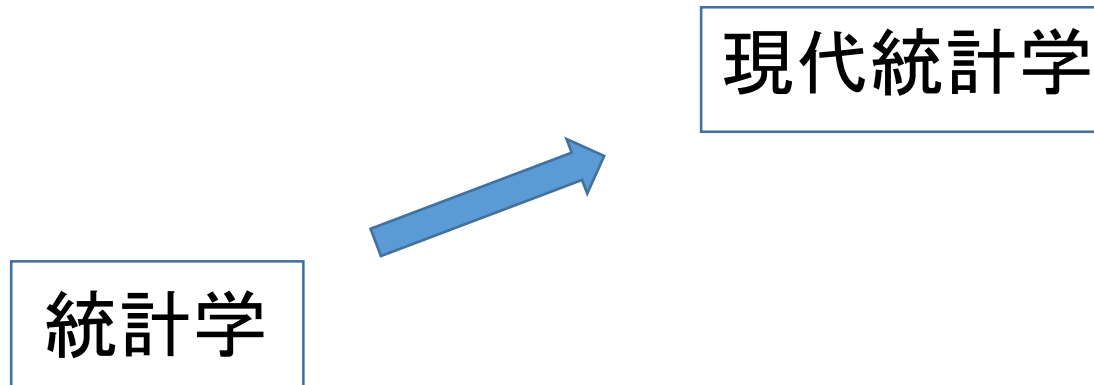


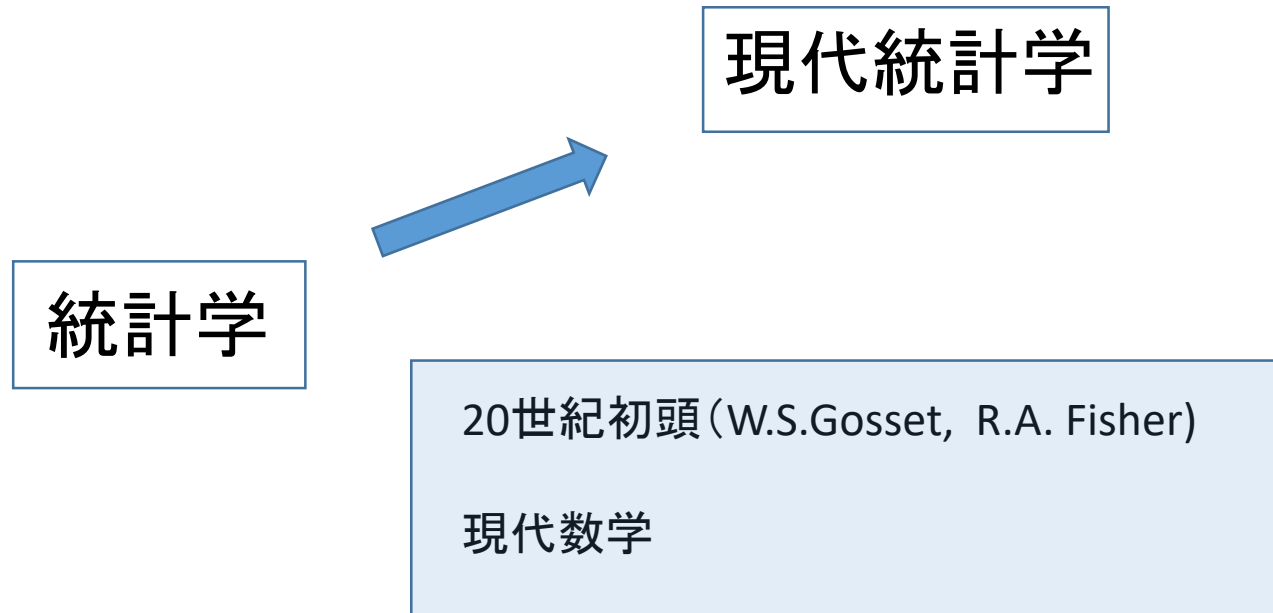
TRADが牽引するデータサイエンス

データサイエンスコンソーシアム, 慶應義塾大学 柴田 里程

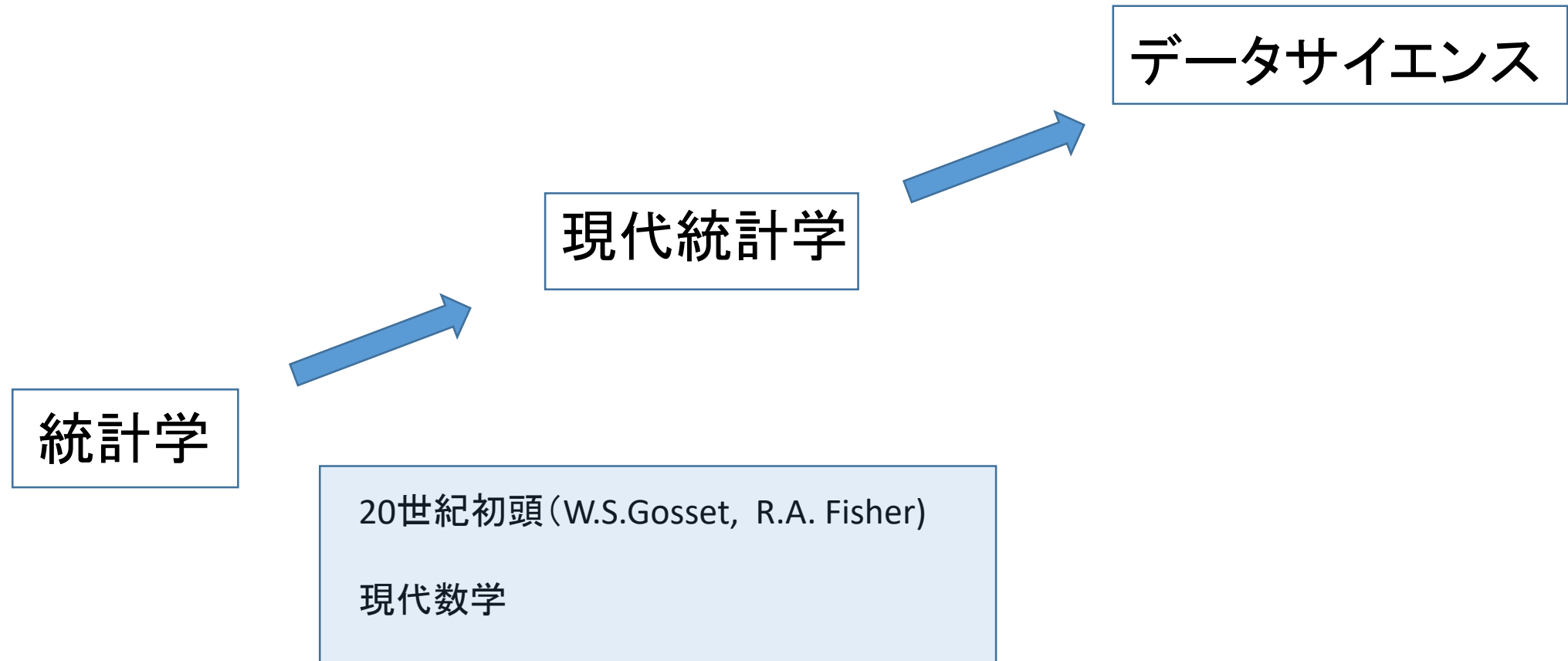
パラダイムシフト(科学革命, クーン)



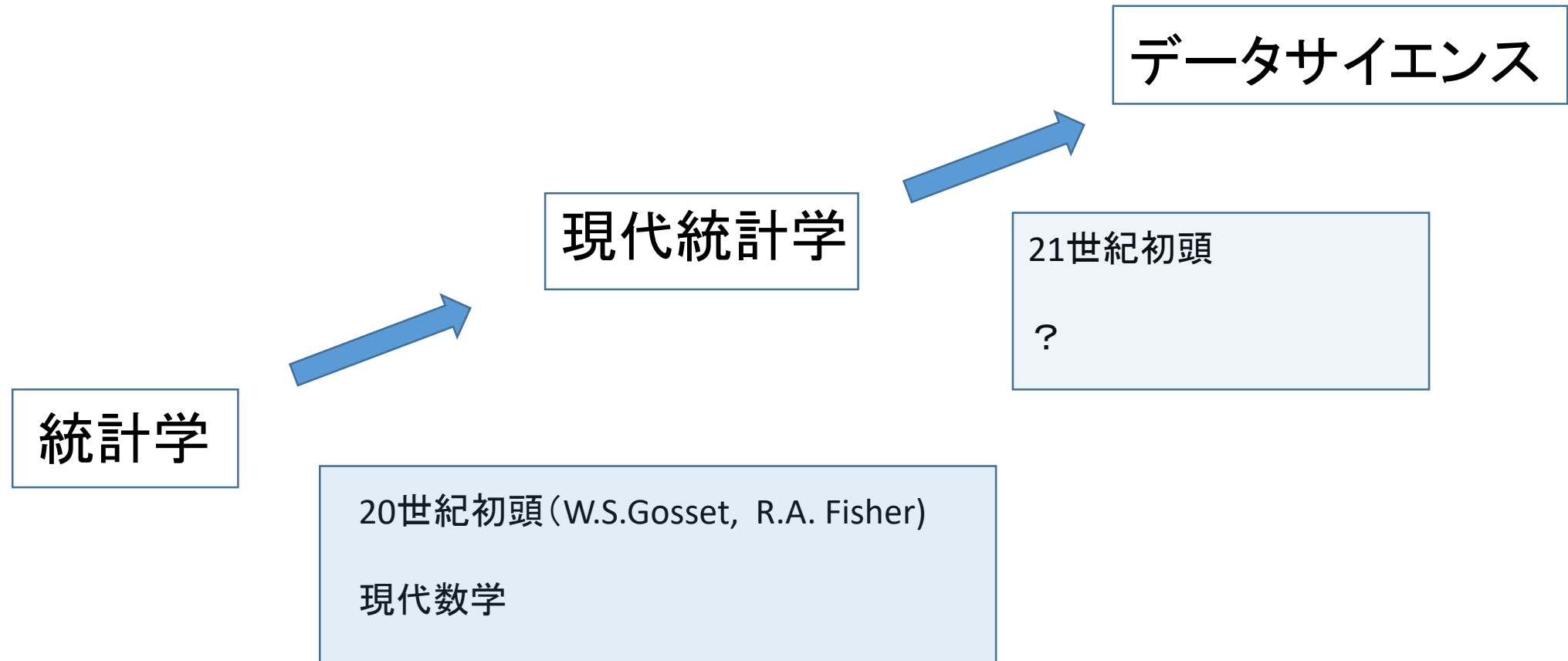
パラダイムシフト(科学革命, クーン)



パラダイムシフト(科学革命, クーン)



パラダイムシフト(科学革命, クーン)



現代数学導入の功罪

- 的確なコミュニケーション
 - 深い思考への入り口
 - 体系化
 - 汎用化
-
- 標本という形での形式化
 - 現代確率論
 - 変数の概念の明確化
 - データの多様性の捨象



現代統計学の脆弱性

今回のパラダイムシフト

- データの多様性
 - 知見の体系的な蓄積
 - 汎化
- データの質と量
 - 質の評価
 - 量の評価
- データの活用
 - 戦略
 - 深化

今回のパラダイムシフトのキー

一つの成熟段階を迎えた
コンピュータサイエンス

1. 研究者間の的確なコミュニケーション
2. 多様性の体系的な蓄積と汎化
3. データの質と量の評価
4. データの活用
5. データサイエンスの地位確立

コンピュータサイエンスの何を取り入れるか

- 関係形式データベース論
 - データの汎用な扱い
 - 効率的な処理が目的
 - 変数の概念はない,
- オブジェクト指向プログラミング
 - 対象の抽象的な
 - 視点, 視野, 視覚
- 計算アルゴリズム
 - 高度なユーザインタフェース
 - 複雑で大量なデータ

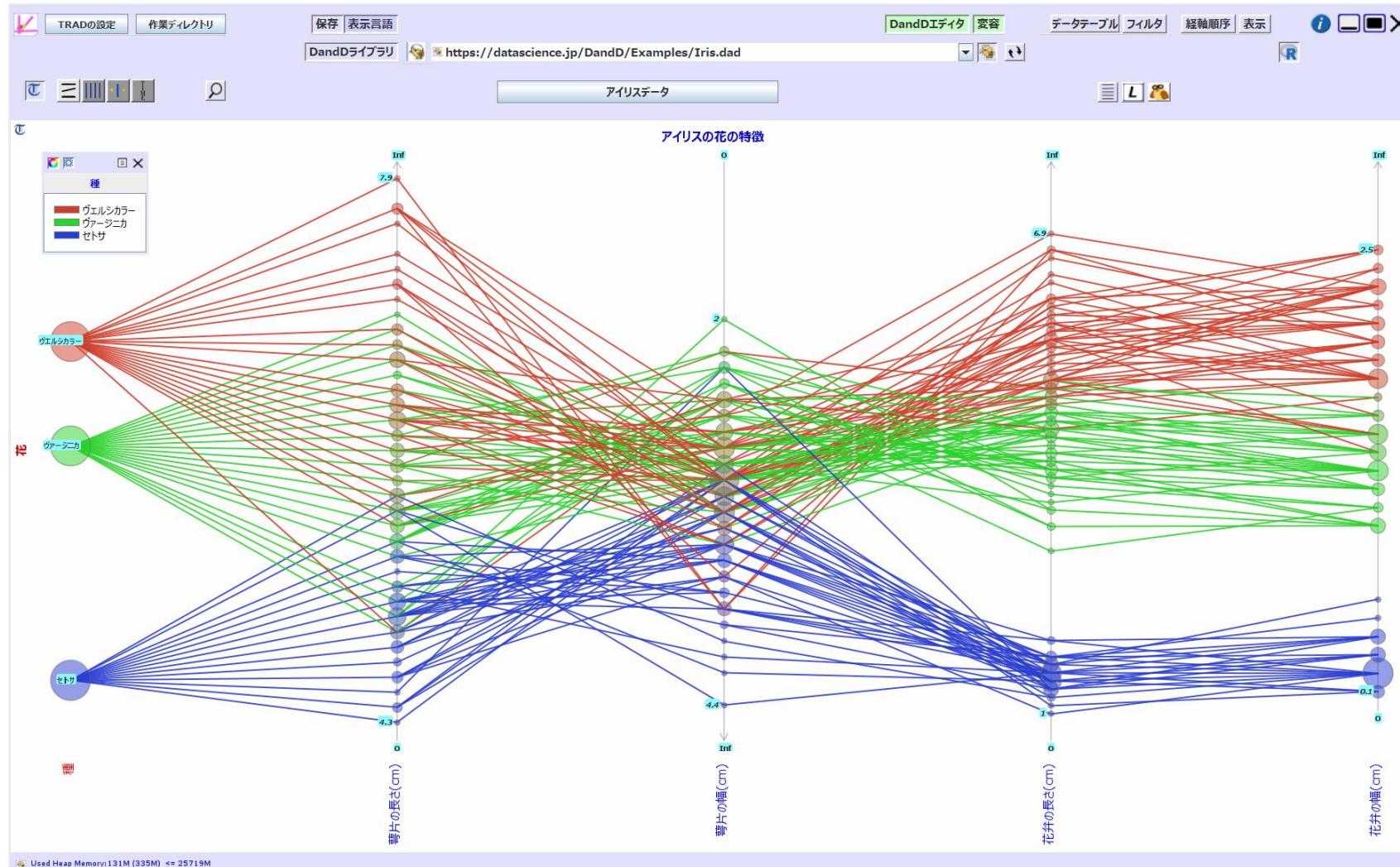
データサイエンス環境

- 数学と並ぶ, データサイエンスを牽引するエンジン
 - S -> R: データ解析環境; プログラミング環境
- 現代的なヒューマンインタフェース
 - GUI (アイコン, マウスクリック)
 - データを中心に据えた視覚的な対話
- 楽しい
- 汎用
 - OSから独立
 - どんな用途にも対応
- 非商用
 - オープンソース
- 多言語対応

TRAD(TextilePlot, R and DandD)

2クリック

TAD



R への GUI

R

The screenshot displays the DandD GUI interface for R. The main window is titled "アイリスデータ" (Iris Data) and shows a table of data for 150 Iris flowers. The table has 5 columns: 種 (Species), 萼の長さ (Sepal Length), 萼の幅 (Sepal Width), 花弁の長さ (Petal Length), and 花弁の幅 (Petal Width). The data is sorted by species, with 127 flowers of the same species at the top and 150 flowers of the same species at the bottom.

Below the table, there is a section titled "作業ディレクトリ: C:\Users\shibata\Documents (TRAD.RData 読み込み, バックアップ/120秒)" (Working Directory: C:\Users\shibata\Documents (TRAD.RData loaded, backup/120s)). This section contains a list of R objects and their classes:

オブジェクト名	クラス
a	matrix, array
invCoefs	function
iris	data.frame
iris2	data.frame

On the right side of the interface, there is a vertical plot titled "花弁の幅(cm)" (Petal Width (cm)). The plot shows a series of green circles representing the petal width of each flower, with a red line connecting the points. The y-axis ranges from 0 to 2.5 cm.

At the bottom of the interface, there is a code editor window titled "transferFun" containing the following R code:

```
1 function(a, n=c(20,15)){
2
3     # a: a[k1, k2]
4
5     dnames=dimnames(a)
6     if(is.null(dnames)) return(NULL)
7
8     k=list(as.integer(dnames[[1]]), as.integer(dnames[[2]]))
9     j=list(0:(n[1]-1), 0:(n[2]-1))
10
11     jk= outer(outer(j[[1]]/n[1],k[[1]]), outer(j[[2]]/n[2],k[[2]]), "+") # j1*k1+j2*k2
12     jk=aperm(jk, c(1,3,2,4)) # (j1, k1, j2, k2) -> (j1, j2, k1, k2)
13     dim(jk) =c(length(j[[1]])*length(j[[2]]), length(k[[1]])*length(k[[2]])) #to be a matrix
14
15     z=exp( 2*pi*(0+1i)*jk )
16
17     f= z %*% c(a)
18
19 }
```

The bottom status bar shows "Used Heap Memory: 131M (335M) <= 25719M".

R GUI の特徴

- TAD とのシームレスな連携
 - データテーブル ↔ データフレーム
- R オブジェクトパネル
 - クラスとともに作業スペースのオブジェクトをリストアップ
 - オブジェクトの表示, 削除, 名前変更, 保存, エディタ起動
- R オブジェクトエディタ
 - いくつでもエディタを開いておける
 - 構文チェック機能を内包
- 作業スペースの R オブジェクトの自動保存
 - デフォルトで 120 秒おき
- 日本語の完全な対応
 - Noto フォント

困難

- OS x 多言語 x R x Java x パブリック
 - OS: 異なるOS (Windows, Mac, MacM1) への対応
 - 多言語: 辞書の更新, ヘルプの翻訳
 - R: 毎年行われるメジャーアップデートへの対応
 - Java:
 - openJDK へ移行し JRE を付属させることでかなり楽に
 - 隠れたバグへの対応(回避)
- パブリック
 - OS や Anti Virus ソフトウェアのセキュリティ強化への対処
 - Signing, Normalization
 - ローカルな HTML ファイルのブラウジング禁止
 - ダウンロード
 - ファイアーウォール
 - 組織によっては一切禁止