

# ジャパンサーチのシステム・ アーキテクチャに関する議論

大向 一輝

東京大学大学院人文社会系研究科・准教授

国立情報学研究所・客員准教授

2019年11月6日

# 本日の内容



[バックナンバー](#) [イベントカレンダー](#) [レ](#)

ACADEMIC RESOURCE GUIDE (ARG) は、1998年に創刊されたインターンです。

## 743号 (2019-04-29、4154部)

投稿者：arg 投稿日時：2019/04/29 (月) 10:39

ACADEMIC RESOURCE GUIDE (ARG)の743号 (2019-04-29、4154部) を発行しました。

主な記事は、

○羅針盤「ジャパンサーチのシステム・アーキテクチャ」 (川島隆徳)

<http://www.arg.ne.jp/node/9738>



[バックナンバー](#) [イベントカレンダー](#) [レ](#)

ACADEMIC RESOURCE GUIDE (ARG) は、1998年に創刊されたインターンです。

## 764号 (2019-09-23、4100部)

投稿者：arg 投稿日時：2019/09/23 (月) 08:17

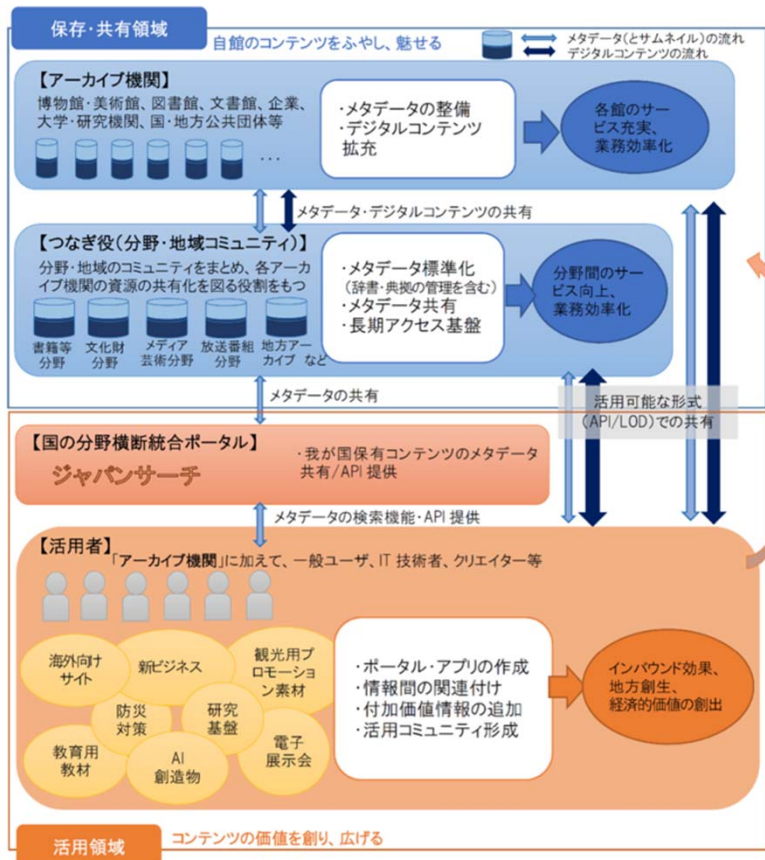
ACADEMIC RESOURCE GUIDE (ARG)の764号 (2019-09-23、4100部) を発行しました。

主な記事は、

○羅針盤「『ジャパンサーチのシステム・アーキテクチャ』を読む」 (大向一輝)

<http://www.arg.ne.jp/node/9906>

# 論考の趣旨



出典：『第二次中間取りまとめ』

<https://jpsearch.go.jp>

# 「決めない」

- 一般論としてシステム開発は決断の連続
  - 決断 = 仕様を確定・固定すること
    - Point of no return
    - 決めると進む。戻ると多大なコストを支払うことに。
  - 未知の状況・不確定要素の中で決断するのがリーダーの仕事？
  - とくにシステム外部に対する仕様はクリティカル
- 決断の先送り
  - ただの先送りは百害あって一利なし
  - アジャイル = 決断のタイミングを遅らせるメカニズム
- 決断を不要にする
  - アーキテクチャの設計によって
  - 新しい技術によって

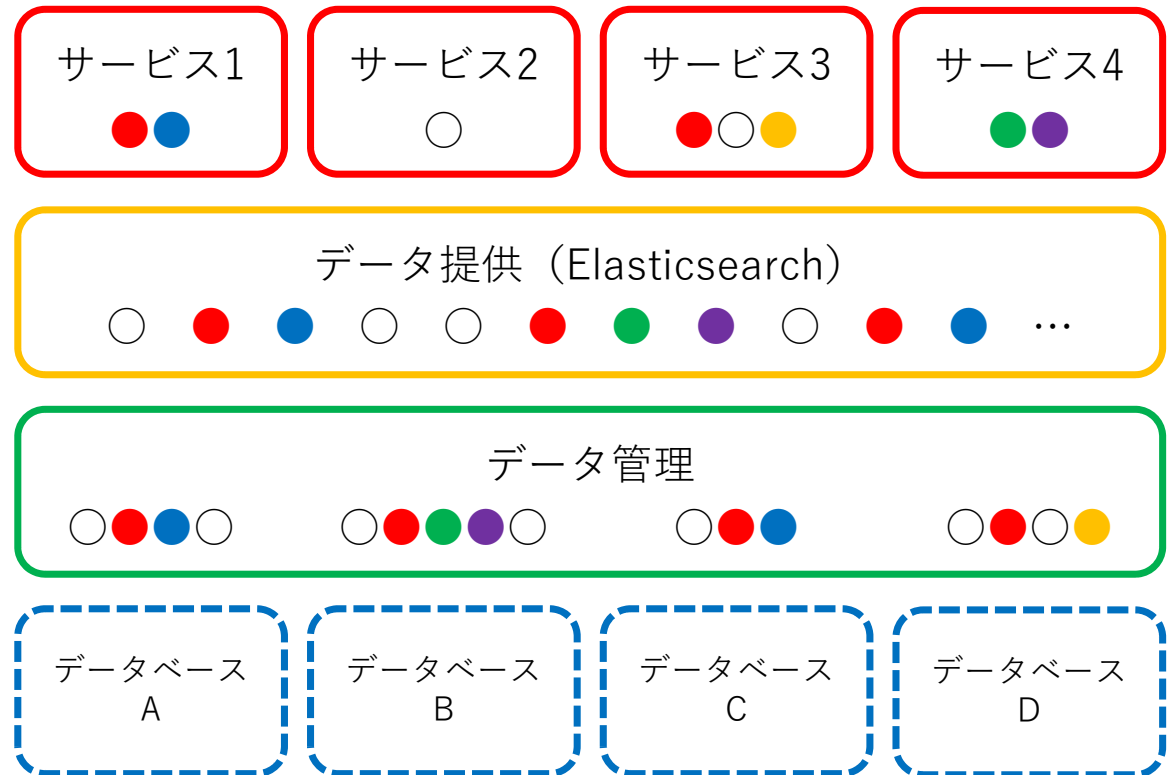
# 「決めない」アーキテクチャ

- 何も決めないと何も作れないが…
  - 入口：データ提供元のメタデータフォーマットは多様。ただし無限の自由度があるわけではない。
  - 出口：サービスへの要求の可能性は無限
- データ管理
  - 緩いコントロールをかける
    - IDと名称（とリンク）のみ必須
    - 共通項目があると望ましい
    - その他は来たものを受け入れる
- データ提供
  - 可能な限り自由度を確保する
    - リクエストの種類によって多様な振る舞いをする
    - 過剰な作り込み（カスタマイズ）をしない
  - サービスの機能 = リクエストのプリセット
    - いくらでも増やせる。あとはユーザインタフェースをつけるだけ。

# ジャパンサーチのアーキテクチャ

## • 3層構造

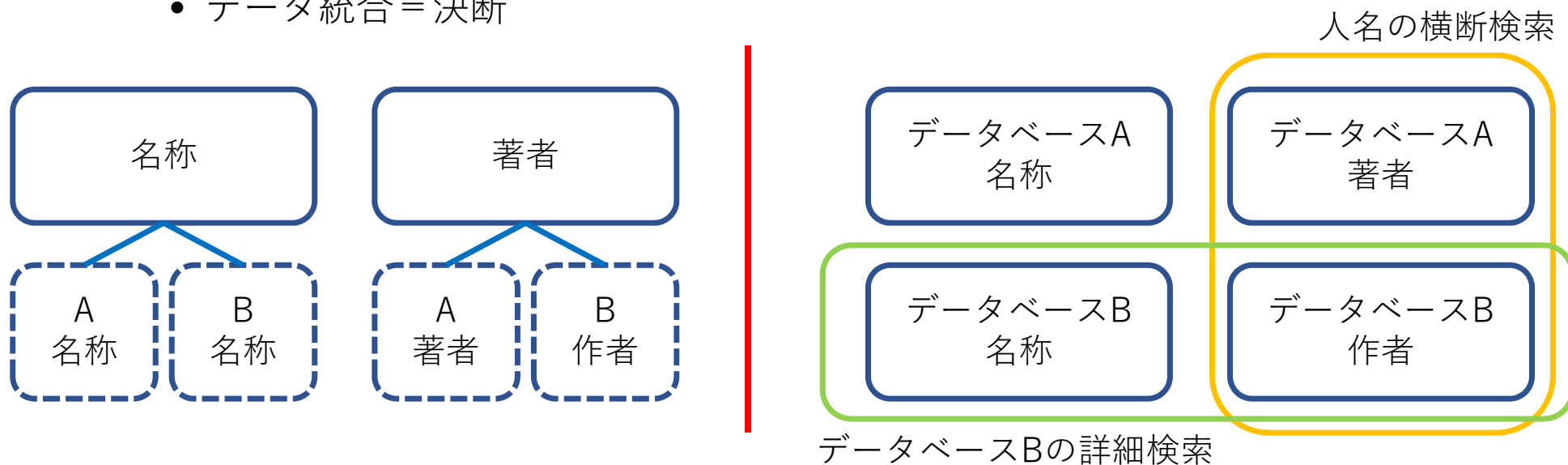
- サービス提供レイヤー
  - 情報の表示
  - UI
- データ提供レイヤー
  - 検索
  - リスト
- データ管理レイヤー
  - 保存
  - 変換
  - マッピング



# データ提供レイヤー

- Elasticsearchの採用

- オープンソースの検索エンジン ex. Solr
- 検索時に検索対象（インデックス）を複数個指定できる
  - インデックス作成のために事前のデータ統合を行わなくてもよい
  - データ統合 = 決断



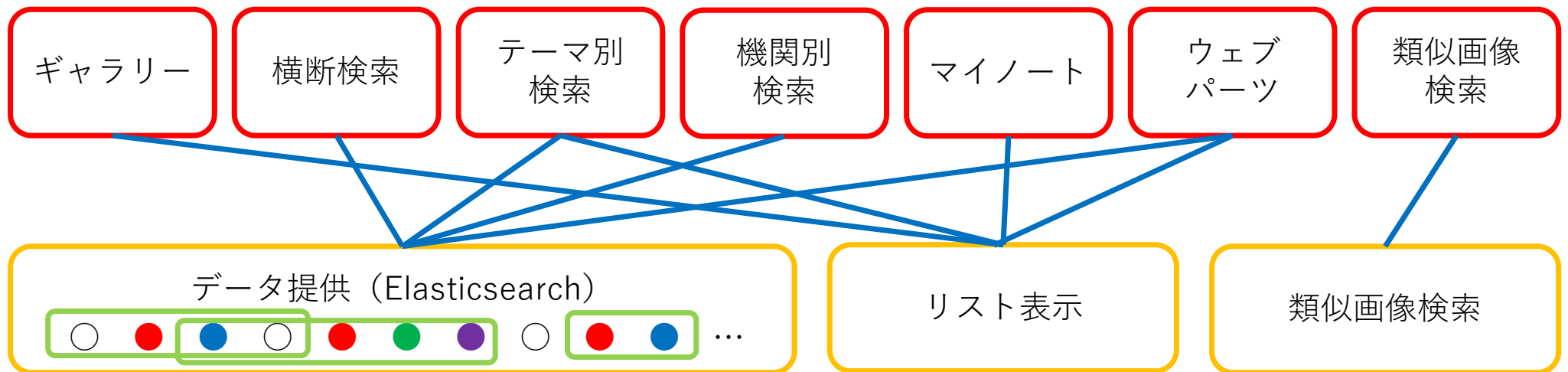
# データ管理レイヤー

- 保存と利用の分離
  - 保存：非破壊的にすべての情報を保持
  - 利用：提供用に適宜加工
    - 共通項目とのマッピング
      - 手動（のためのユーザ管理とUIの提供）
      - 半自動（データ分析？ヒューリスティクス）
    - 利活用スキーマとのマッピング
      - 値の自動判別（時間・空間・人名）
    - Elasticsearch用のデータへの変換
  - 「デジタルアーカイブの汎用的なプラットフォーム」
    - 最も手がかかっているのはおそらく赤字の部分
    - 後戻りの可能性は低い



# サービス提供レイヤー

- 多彩なサービスメニューは基本機能の組み合わせ
  - 検索時に検索対象・検索条件を指定できる強みを活かす
- ユーザーインターフェースはひとまとまりのモジュールを形にしたもの
  - 横展開・新機能の可能性



# 所感

- 素朴な感想
  - 「作っただけでえらい！」
  - 「動いているのでなおさらえらい！」
  - 「素のElasticsearchにUIをかぶせたもの」
- 未知の要求に対する現時点での回答としてほぼベスト
  - データ管理レイヤーのきめ細やかさ（と大胆さ）
  - データ提供レイヤーにそれほど手が入っていない
    - 今後の大方針転換があったとしても対応可能
  - サービス提供レイヤーがデータ提供レイヤーに忠実
    - 独自の仮説（作り込み）が入っていない
    - 多様な展開と即応性を示唆

# 未だ見ぬサービス

- ジャパンサーチの「真の」要求は自明ではない
  - Europeana Collection・DPLAのコピーが欲しいならともかく…
  - 横断検索・詳細検索・全文検索…
  - 電子展示会・画像検索…
  - など開発者に寄せられる「仮説」の数々
- 要求はユーザが事後的につくる
  - 成功したインターネットサービスの典型的なストーリー
  - 兆候が見つかった時に即応できるかどうか肝心
- よい仮説とは？よくない仮説とは？
  - ユーザに開かれている or 閉じている
  - 大量の決断が必要な仮説。戻れない仮説。

# 課題

- 「決めない」ことのトレードオフ
  - 高速化・最適化は難しい
    - 決めてから着手できることがほとんど
    - まだそのタイミングではないが、データ量、アクセス数が増大すると…？
    - サーバの物量で解決できる問題はそちらで
  - 強力なチームが必須
    - 「決めない」はメタメッセージ
    - 意図を共有できる少数精鋭のチームだからこそ実現可能
    - 決めることが前提の外注モデルを取りにくい
  - いつかは「決める」
    - それはいつなのか？
    - 何に決まるのか？