

# 物質・材料研究機構における研究データ 管理に向けた取り組み

## The current approach for Research Data Management at NIMS



国立研究開発法人 物質・材料研究機構  
材料データプラットフォームセンター  
菊地伸治

## ➤ 「第四の科学手法」の提唱から既に10年を経過

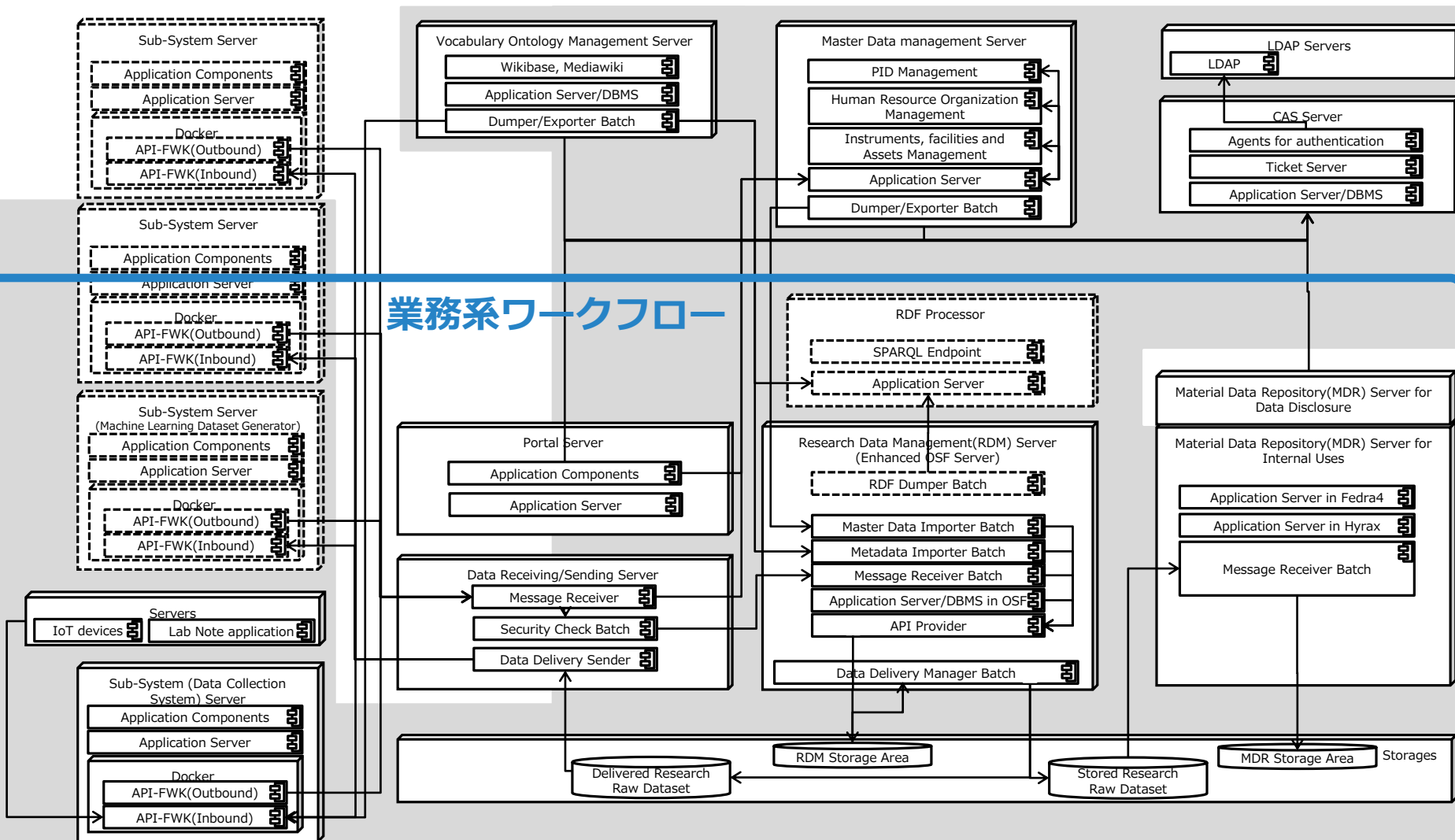
- ✓ The Fourth Paradigm: Data-Intensive Scientific Discovery
- ✓ **[出典]** T. Hey, S. Tansley and K. Tolle Published by Microsoft Research, October 2009, ISBN: 978-0-9825442-0-4
- ✓ <https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/?from=http%3A%2F%2Fresearch.microsoft.com%2Fen-us%2Fcollaboration%2Ffourthparadigm%2F>
- ✓ クラウドコンピューティング、センサーNW、機械学習技術が定着・高度化

## ➤ 材料科学分野でも世界規模でMaterial Informaticsに関する取り組みが進行中

- ✓ 蓄積された膨大な実験データ、計算機能力の向上により算出可能となった膨大な計算データを入力として統計学、パターン認識、AI等のデータ解析技法を用いてプロセスと特性間、異なる特性間に成り立つ法則性を抽出・発見、予想を可能とすることで、新たな材料開発を加速すること。
- ✓ **[出典]**
  - ✓ 知京豊裕, 'マテリアルインフォーマティクスの現状と課題～海外の動向と日本の挑戦', 情報知識学会誌, 2017 Vol.27, No.4, Pp.297-304, 2017.
  - ✓ 上島伸文, 及川勝成 '技術解説～計算材料科学・工学の最新動向', 電気製鋼, 第87巻1号, 2016年, Pp.21-26.

## ➤ 当機構でも、材料科学に纏わる各種データを「つくる」「ためる」「使う」「公開する」という4機能を統合したMaterial Informatics環境実現に向けて公共財としての「材料データプラットフォーム」を整備中



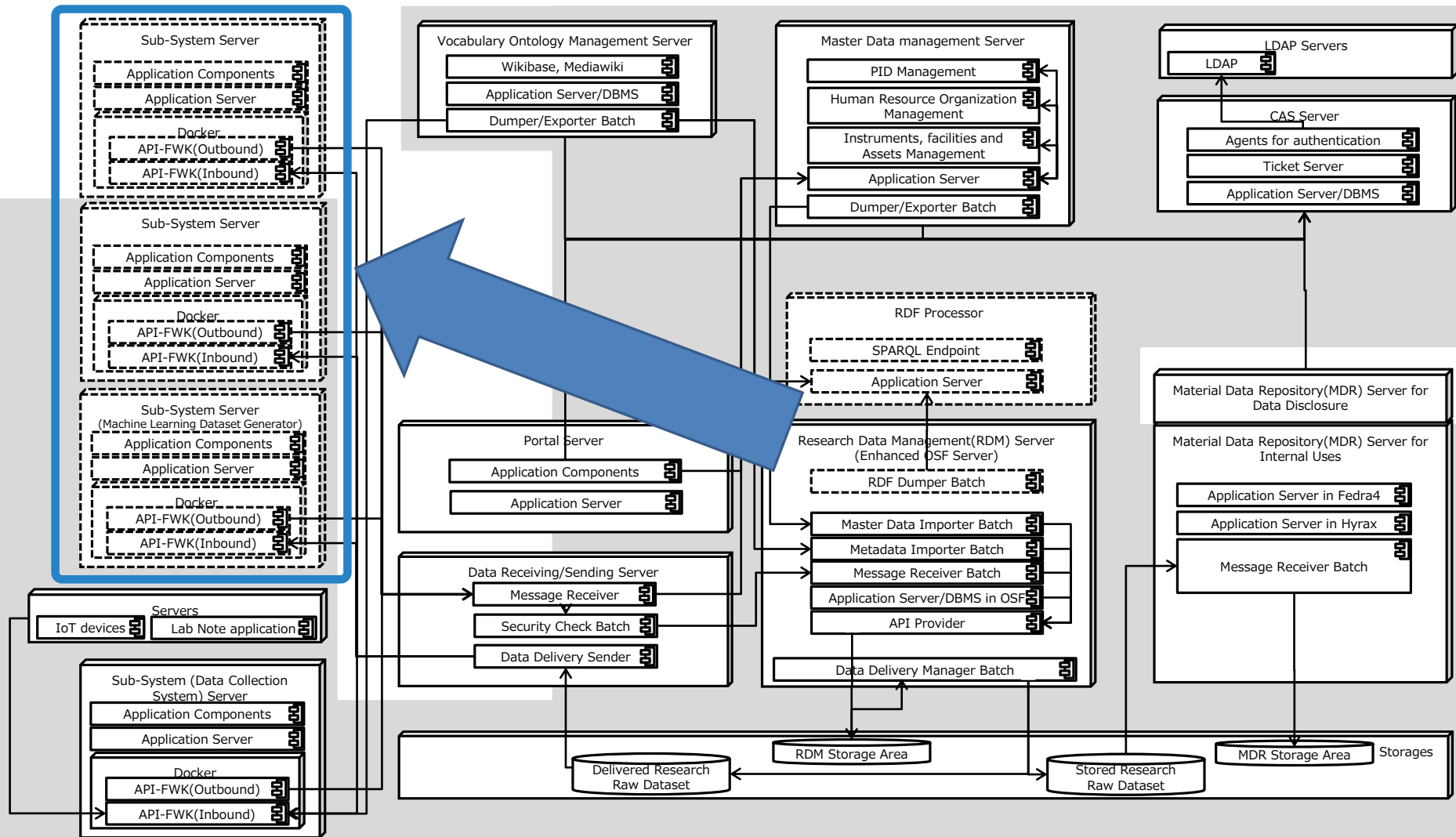


生成する(つくる) →

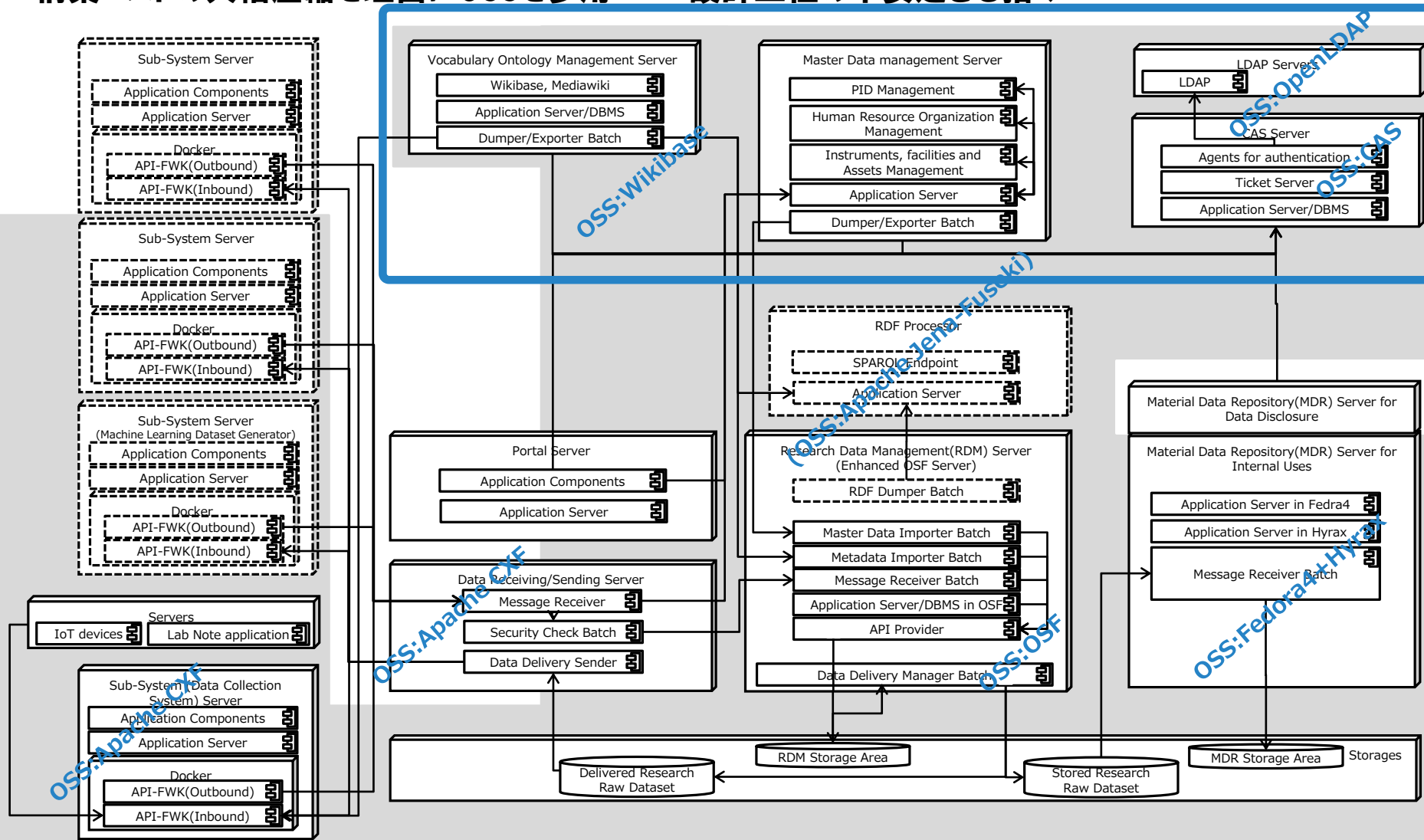
管理する(ためる) →

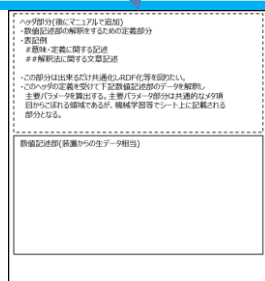
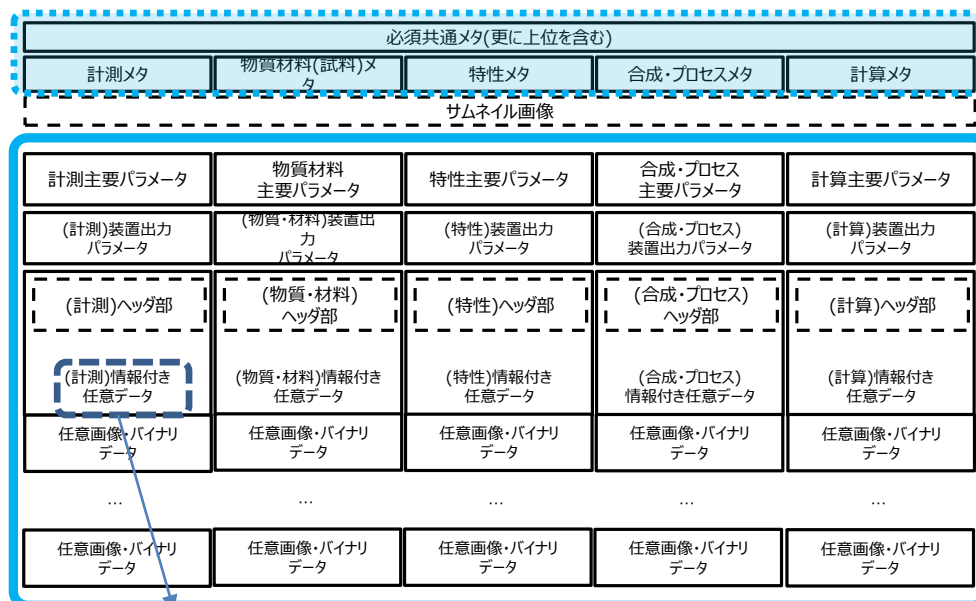
公開する

## 利活用する(つかう)

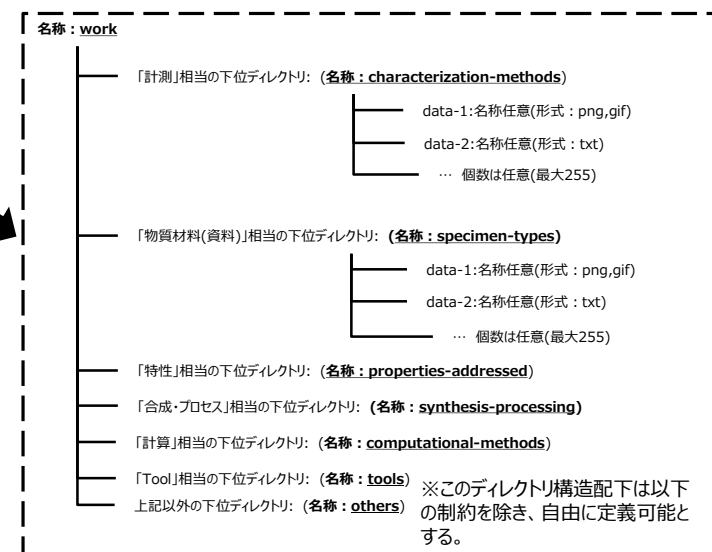
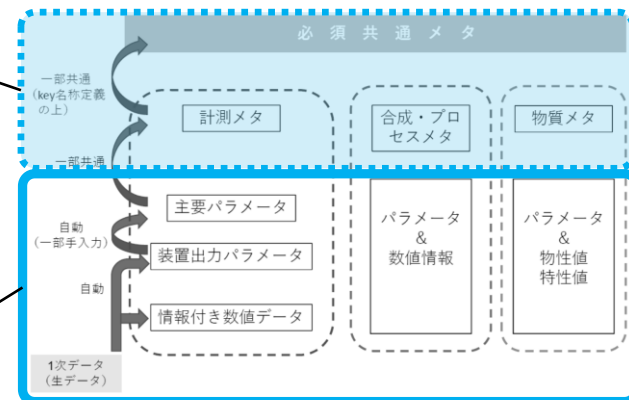


➤ 構築コストの大幅圧縮を理由にOSSを多用 → 設計工程の不安定さも招く

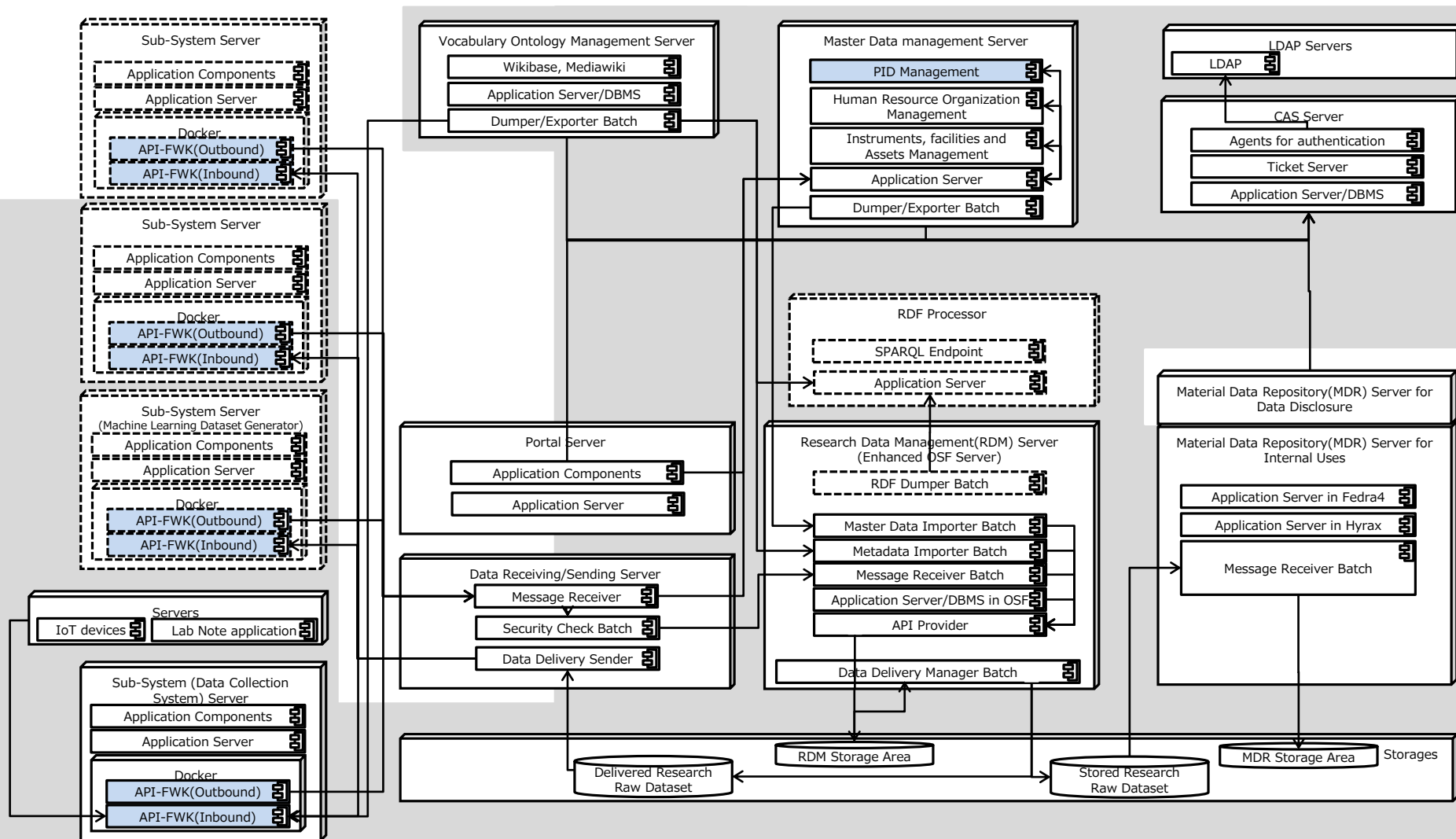




アーカイブで統合・圧縮の後、Base64でエンコード・暗号化される。その後、全体でpidを付番し、署名付与をしてメタデータ(Json)と共に1メッセージで登録を実施する。



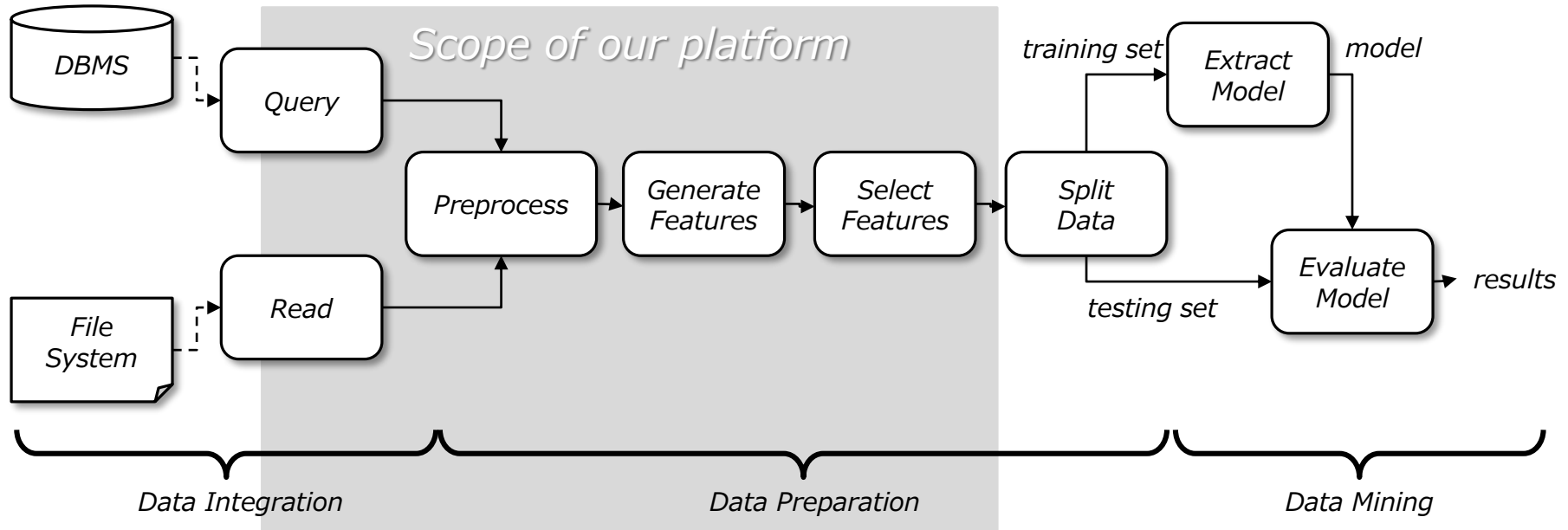
## ➤ API-FWKは後述



## ➤ PID (Persistent Identifier) : Master Data Managementの一環、試料・装置等に総背番号化、データの意味的統合時のGlueの役割



## ➤ 役割的には前工程が中心、但し、意味的な統合迄の実現を指向

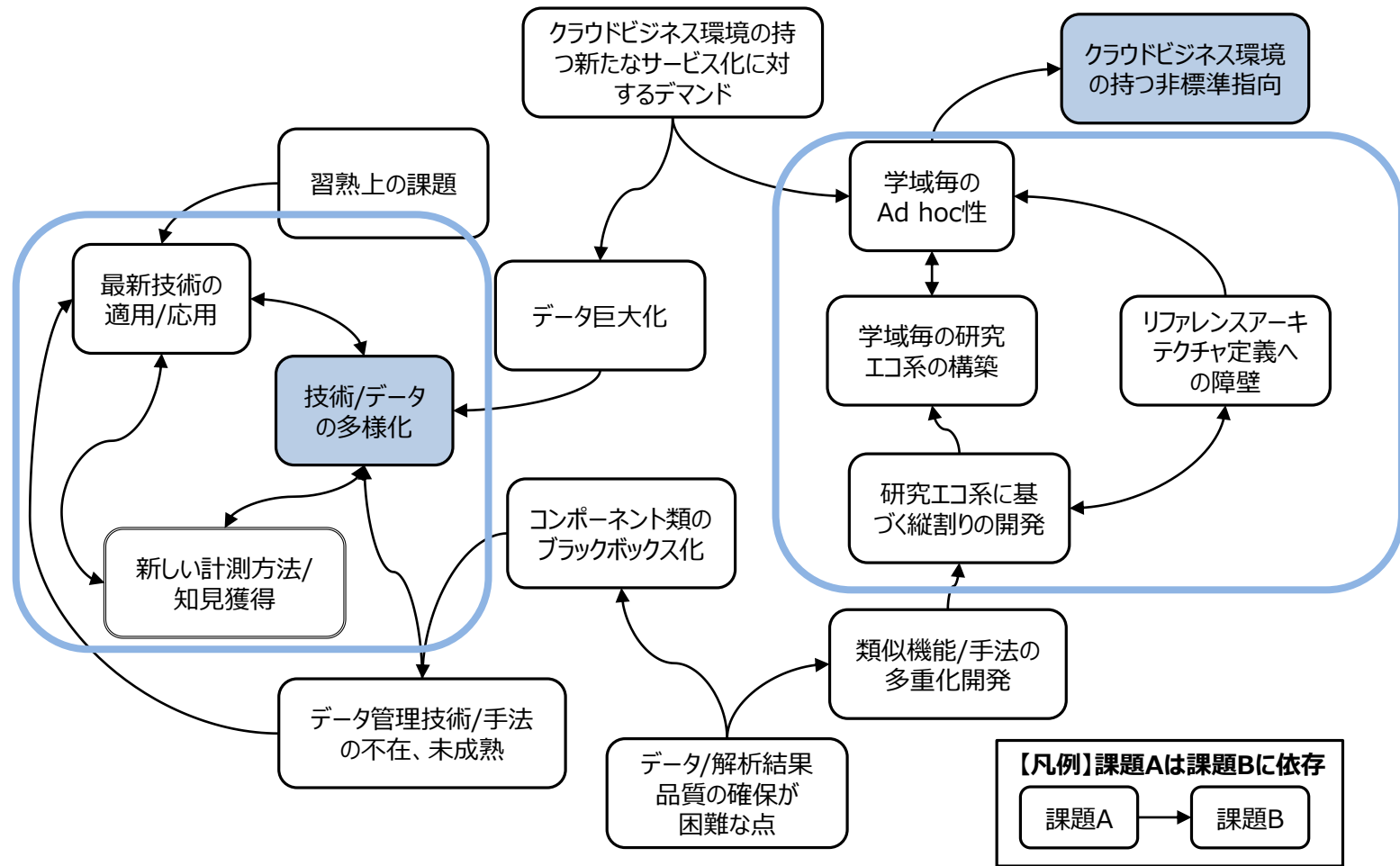


【出典】 C.S. Liew, et al. 'Scientific Workflows: Moving Across Paradigms'. ACM Computing Surveys, Vol. 49, No. 4, Article 66, 2016.

## ➤ 逡増的に基本設計を進化させており「マイクロサービス」の性格がより高まっている

✓ 構築上はネガティブな面も含むが、「進化的アーキテクチャ」の概念を取り込みつつある

【出典】 N.Ford, R.Parsons and P.Kua. 'Building Evolutionary Architectures – Support Constant Change - O'Reilly Media, Inc, 2017



[出典] 菊地伸治, 細野繁, 'Scientific Workflowの動向について' 信学技報, vol. 118, no. 72, SC2018-5, pp. 27-31, 2018年6月.

- 有向アークを集めるノード、有向アークの終着ノードが、真因の可能性が高い
- 閉じたループを作るノード群は構造的な問題であり、解決が難しい

- 前課題構造で見られる下記課題群に対する見解は以下の通り。特に「マイクロサービス」指向を含めた「進化的アーキテクチャ」への展望により、部分的にせよ潜在的には対応し得ると考える
  - 技術/データの多様化：流通管理と解析部を分離・RDF化、管理機構は標準化
  - クラウドビジネス環境の持つ非標準指向：未対応事項、OSF-Waterbutler API採用等で対応
- 意味的な統合迄の実現を指向しており、その意味で「高付加価値科学データ創出を指向した研究データ管理プラットフォーム」と定義し得る。
  - 更にリファレンスモデルに昇華させるには、「進化の取り込みやすさ」を考慮したブラッシュアップも必要。