

第1回SPARC Japanセミナー2020

「研究データ公開：フルオープンと制限公開の境界線」登壇資料

データ流通市場における データ公開の事例・論点

すべての課題解決に役立つ、あらゆるデータをお届けします。

2020年10月2日

株式会社日本データ取引所 上島邦彦

1



会社概要



日本データ取引所 (Jdex)

- データ取引専門会社として創業5年目
 - 社内データ活用計画の立案
 - データ調達の仲介
 - データ製品の企画・設計
 - 組織間の法的相互運用性の向上
- ガイドライン策定・国際標準化活動にも貢献
 - データカタログ
 - データ品質
 - データライセンス

出所：「兼松、データ取引専門企業の株式会社日本データ取引所と資本業務提携」（2019）をもとに作成



本日の話題

「鍵のかかった部屋」をいかに持ち運ぶか



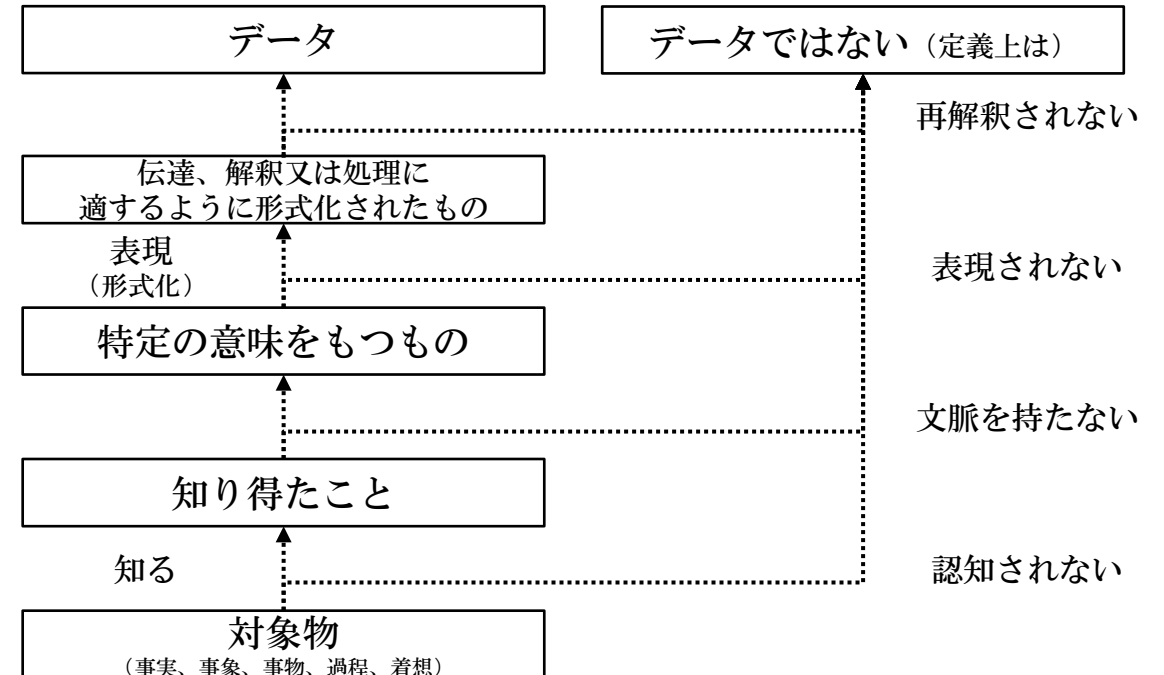
用語の説明

データとは何か

ざっくりいうと

- **記録**された事実
- 価値ある情報の**素材**
- **さまざまな形態**
(数値、図表、統計、文章...etc.)

厳密に定義すると

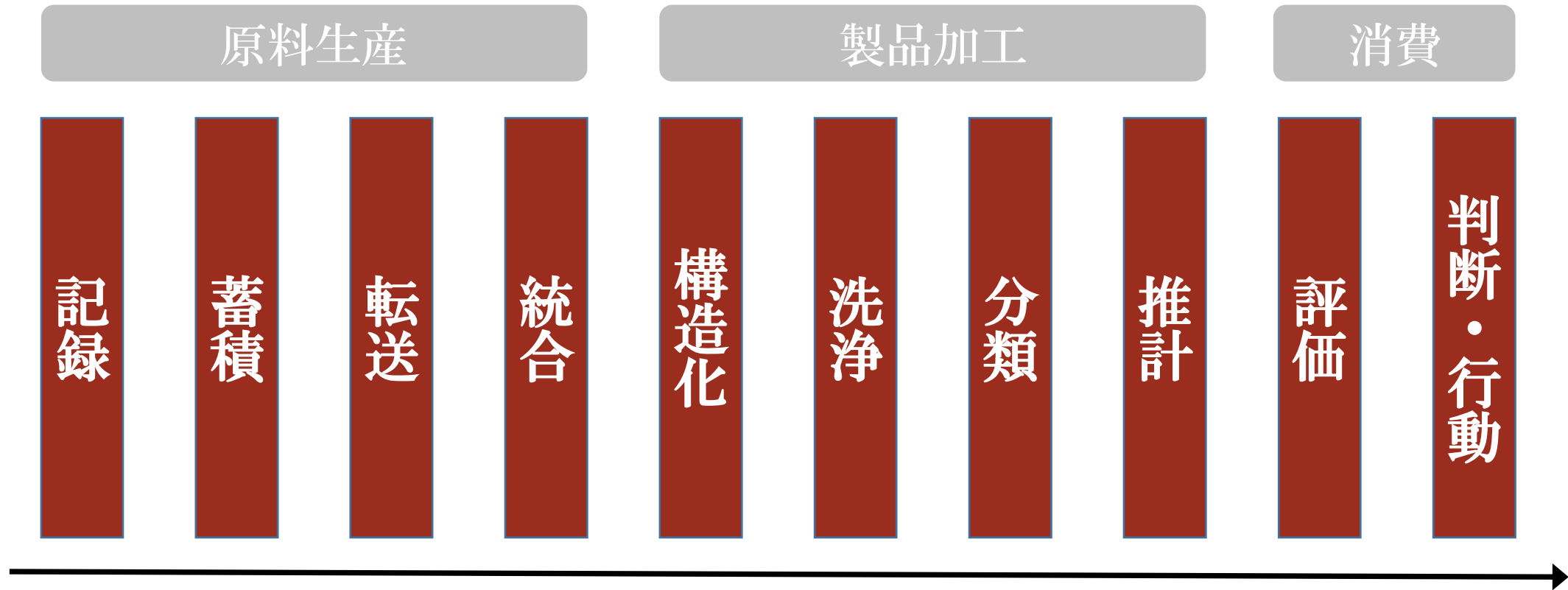


- データとは、情報の表現であって、伝達、解釈又は処理に適するように形式化され、再度情報として解釈できるもの
- 情報とは、事実、事象、事物、過程、着想などの対象物に関して知り得たことであって概念を含み、一定の文脈の中で特定の意味をもつもの

出所：日本工業規格「基本用語x00001-1994 01.01.01」

データ流通の流れ

- データは多くの工程を経て流通し、ビジネスや学問、行政などを支えています。
- 他の多くの製品・コンテンツと同じく、原料を生産し、製品に加工し、消費されます。



データ製品という考え方

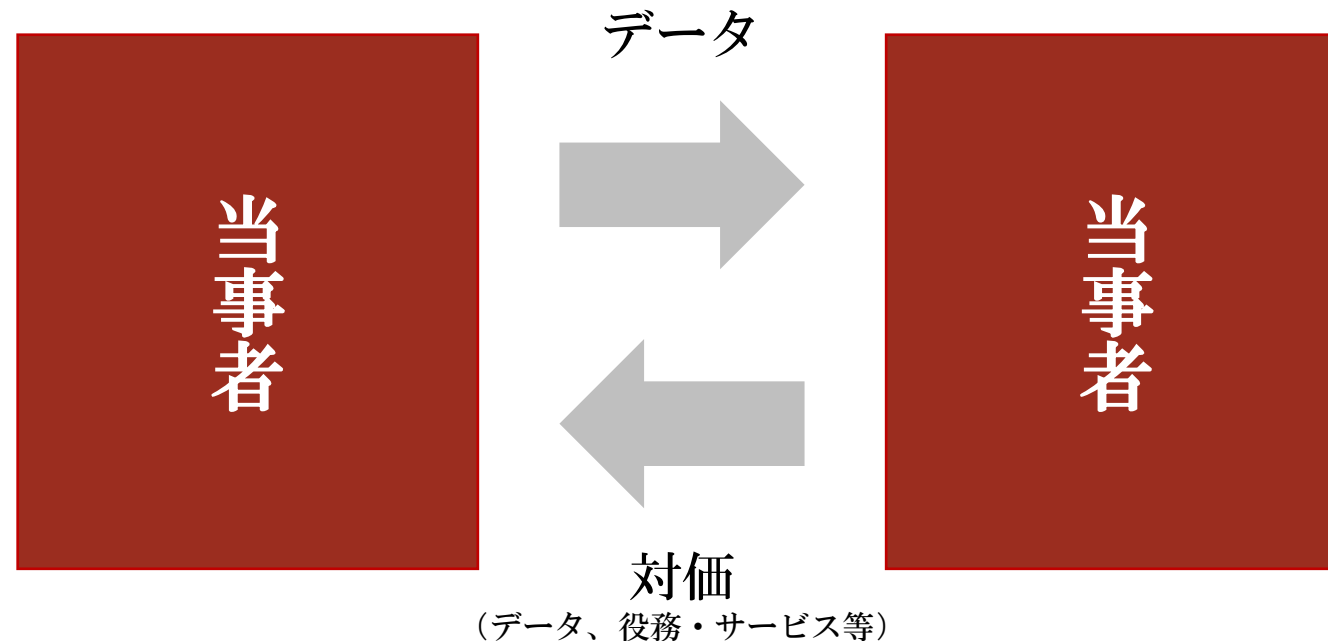
データ製品とは、データセット、統計、図表・グラフ、報告書など、生の情報を加工して作られるデジタルコンテンツの総称です。

その製品情報を要約・整理し、一覧にしたものを**データ製品カタログ**と呼びます。

※当社や取引先のあいだで通用している、世間にはまだまだあまり普及していない新語です。
※学術研究における「出版対象データセット」「研究データリポジトリ」とほぼ同義でしょうか。

データ取引とは何か

- 誰かが持っているデータを、他の相手に見せたり、渡したりすることを指します。
- 対価が金銭とは限らず、互いにデータを開示したり（情報交換）、役務・サービスが提供されることもありますし、対価のない取引（例：贈与、譲渡、寄付）もあります。



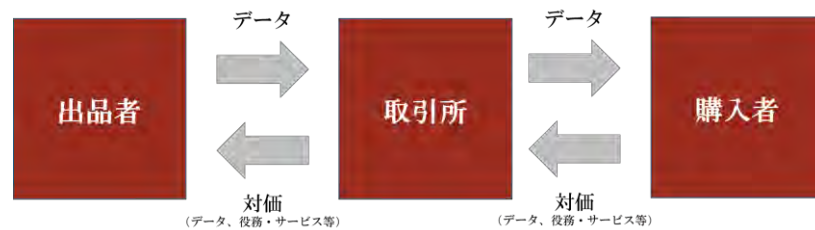
※本発表では、研究データを共有した共同研究も、広義のデータ取引と見なしています。

データ取引所とは何か

※本題ではないので詳細は割愛。

- 諸説ありますが、「取引とは何か」の考え方によって、大きく4通りに分けられます。

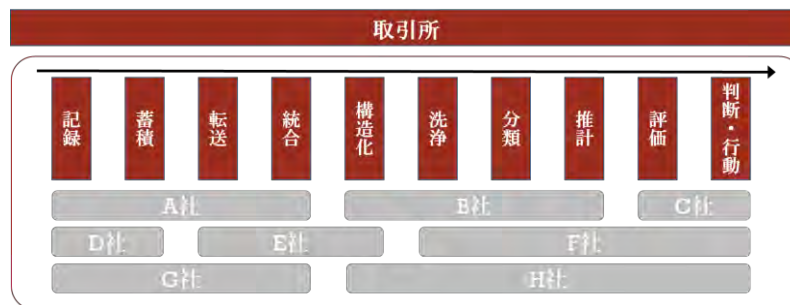
1.代理店型



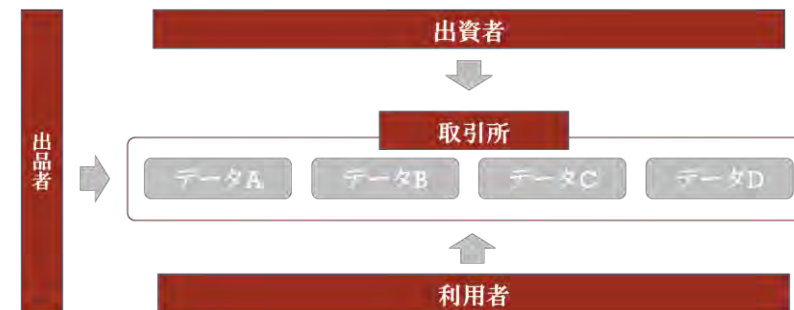
2.市場型



3.共同組合理型



4.保管所型





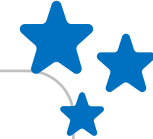
本日の話題

「鍵のかかった部屋」をいかに持ち運ぶか

データ公開の方法整理 (イメージ)

- 情報サービス・出版事業を専らとしない私企業にとって、フルオープンなデータ公開は遠くの夢……。

フルオープンの世界



不特定

利用者限定 (制限共有?)

- ・ 専門メディア掲載
- ・ ソースコード共有
- ・ データリポジトリ
- ・ データ検索サービス
- ・ 統計・報告書の販売
- ・ 業界統計・年鑑

一般公開 (制限公開?)

- ・ デジタルコンテンツ
- ・ セミナー・講演
- ・ 出版物
- ・ プレスリリース
- ・ IR資料
- ・ ホワイトレポート

社内の情報交換 (制限共有?)

- ・ 社内勉強会
- ・ 社内ナレッジ共有
- ・ 社内広報
- ・ グループ間受発注
- ・ 部署間の予算調整
- ・ 私的な貸し借り

取引先に開示 (制限共有?)

- ・ 秘密保持義務
- ・ 業務資料の提供
- ・ データ交換
- ・ 共同事業・サービス
- ・ 共同利用・共同調査
- ・ 人材交流

特定



企業のデータ取扱担当者

少数

多数

企業のデータ公開にありがちな障壁 (1/2)

法的課題

- 国際レベル：通商協定、越境データ移転、ローカリゼーション規制
- 法令レベル：営業秘密、知的財産権、個人情報、パブリシティ権
- 企業レベル：秘密保持契約、データ利用規約、情報管理指針、社内決裁
- 実務レベル：利用条件の整理、契約書の起草、ライセンスの設定

技術的問題

- 統計調査・データ分析と専門分野の双方に精通した人材は少数派（昔も今も将来も）
 - 「欲しいデータ」「外に出していいデータ」の判断も意外にむずかしい
- データ製品開発には多様な技術が介在する（慎重を期すほど高コスト体質に）
 - デバイス性能理解、データ品質評価、データ洗浄手順、匿名化・曖昧化
- 安全なデータ移転のためのセキュリティ確保、プライバシー保護、証跡管理
 - 業務フロー自体に根ざすリスクを減らすには、働き方から変える必要がある

企業のデータ公開にありがちな障壁 (2/2)

・ ブランド毀損

- ・ 業界慣習、社内外の視線、敏感な専門家、無関心な一般生活者の板挟み
 - ・ 炎上リスクに当意即妙で応じられる広報スキルと法務リソースを持つ担当者は少数

・ ビジネス機会の損失リスク

- ・ 「データの保持」を秘すること自体が競争力になる
 - ・ サンプルデータや項目定義だけでも技術流出リスクが懸念
 - ・ サプライチェーンの上流・下流で、交渉材料として使われる情報開示
- ・ 得られる利益（ブランディング、人材獲得、知識習得）の不確実性
 - ・ 経営層・現場ともに「すべきだ」と分かっているにもかかわらず、設備・人員の投資判断を行いつづける
 - ・ 自宅の積ん読がどういうわけか一向に減らないことと似ている

・ ITリテラシー格差

- ・ 組織内のデータ活用スキルのばらつき、データ活用部門の大いなる忙殺
- ・ 取引相手との知識ギャップ（スーツとギークの、理論と応用の、大企業とベンチャーの...）

企業のデータ ≡ 鍵のかかった部屋

目標：だれもが安心して自由に使える

一般論：「鍵のかかった部屋」取扱法

・ 常識通り

- ・ 「部屋」の住人に招き入れてもらう
- ・ 「部屋」の住人になる
- ・ 「鍵」を手に入れる

・ 常識破り

- ・ 「鍵」をなくす（出入り自由にする）
- ・ 「部屋」を解体する（壁をなくす）
- ・ 「部屋」ごと持ち運ぶ（！）

企業データの取扱法になぞらえると

・ 20世紀的方法

- ・ 組織のメンバーシップを入手する（転職、出向、共同研究...etc.）
- ・ データのオーナーシップを持つ（業務提携、権利移転、事業買収）
- ・ **データの利用権限を付与してもらう**（購入申込、利用申請、出版契約）

・ 21世紀的方法

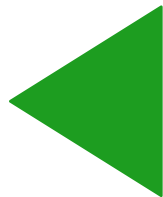
- ・ コミュニティ運営（会員向けサービス、期間限定ハッカソン、リポジトリ開放）
- ・ ライセンス設定（オープンデータ、オープンソース、CCライセンス）
- ・ **マーケットプレイス**（データ製品の出品登録・購入リクエスト）

1. データの利用権限を付与してもらう
2. データ製品の出品登録・購入リクエスト

さまざまな工夫

課題

- ・ 法的課題
- ・ 技術的問題
- ・ ブランド毀損
- ・ 機会損失リスク



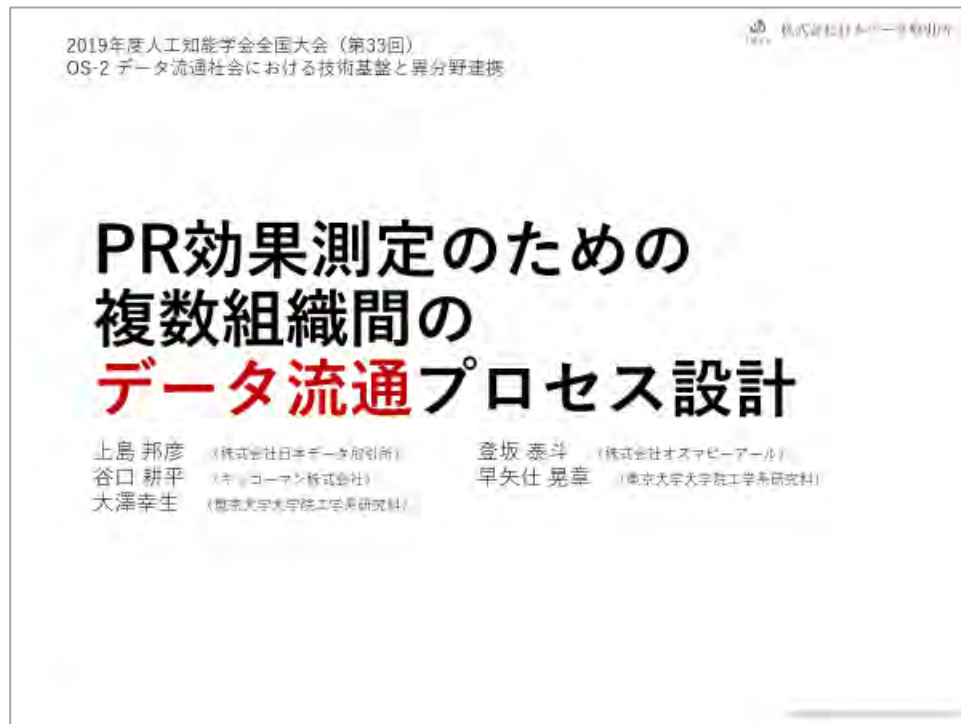
工夫

- ・ 利用規約の作成
- ・ 業務フローの標準化
- ・ UI/UXデザインで事故防止
- ・ 提供内容・方式の棲み分け

※ITリテラシー格差は、時間・労力を惜しまず、詳しい方／疎い方の双方が「認める」「許す」「慣れる」しかありません。

事例：約20社間でデータ流通圏を形成

- 事業会社、PR会社、大学研究室、当社の4社が中心となって行った共同企画。
- データ提供者から「毎年」「個別に」「データ種類ごとに」利用許諾を得て、約20社分のデータを活用可能に。



多種多様な、活用するデータの候補

店頭 <ul style="list-style-type: none"> 監視カメラ, 生体認証 	購入 <ul style="list-style-type: none"> 店舗POS, 決済履歴 	会員登録 <ul style="list-style-type: none"> 購買履歴, メルマガ
テレビ放送 <ul style="list-style-type: none"> 視聴率, 露出秒数 	プリントメディア <ul style="list-style-type: none"> 新聞, 雑誌, 書籍 	屋外広告 <ul style="list-style-type: none"> 交通広告, チラシ, 催事
Webメディア <ul style="list-style-type: none"> ニュース, 動画 	オウンドメディア <ul style="list-style-type: none"> PV・UU・滞在時間 	ユーザー登録 <ul style="list-style-type: none"> オーディエンスデータ
検索エンジン <ul style="list-style-type: none"> 検索キーワード 	検索結果 <ul style="list-style-type: none"> SEO関連指標 	Web広告 <ul style="list-style-type: none"> 行動タグ埋め込み
ソーシャルメディア <ul style="list-style-type: none"> 投稿件数, リーチ数 	SNSコンテンツ <ul style="list-style-type: none"> エンゲージメント 	ID連携 <ul style="list-style-type: none"> 顧客プロファイリング
スマートフォン <ul style="list-style-type: none"> 位置情報 	スマホアプリ <ul style="list-style-type: none"> ダウンロード, 操作 	アプリ内行動 <ul style="list-style-type: none"> 利用, 課金, ユーザ交流



Excel (1枚)

※研究内容に営業秘密や第三者の知的財産権が含まれるため、学会発表資料から一般論および公知の情報のみ抜粋しています。

考察：データの利用権限を付与

- **利点**：やればできる
- **難点**：大変

1. データの利用権限を付与してもらう
2. データ製品の出品登録・購入リクエスト

事例：データ取引システム「JDEX™」

- 売りたいデータを簡単に出品できて、欲しいデータをすぐに探せる会員制のデータマーケットプレイスです。
- オープンデータ、マーケティングデータ、統計データベースなどを扱います（2020年11月登録受付開始予定）。



図：データ検索画面（社内テスト中のダミーデータを表示）

主な特長

- 公開制限のための機能を組み合わせて、参加者ごとの事情に合った対応が行えます。

データカタログ



審査制マーケット



限定注文、限定出品



匿名加工サポート



ライセンス契約自動作成



おすすめデータ紹介

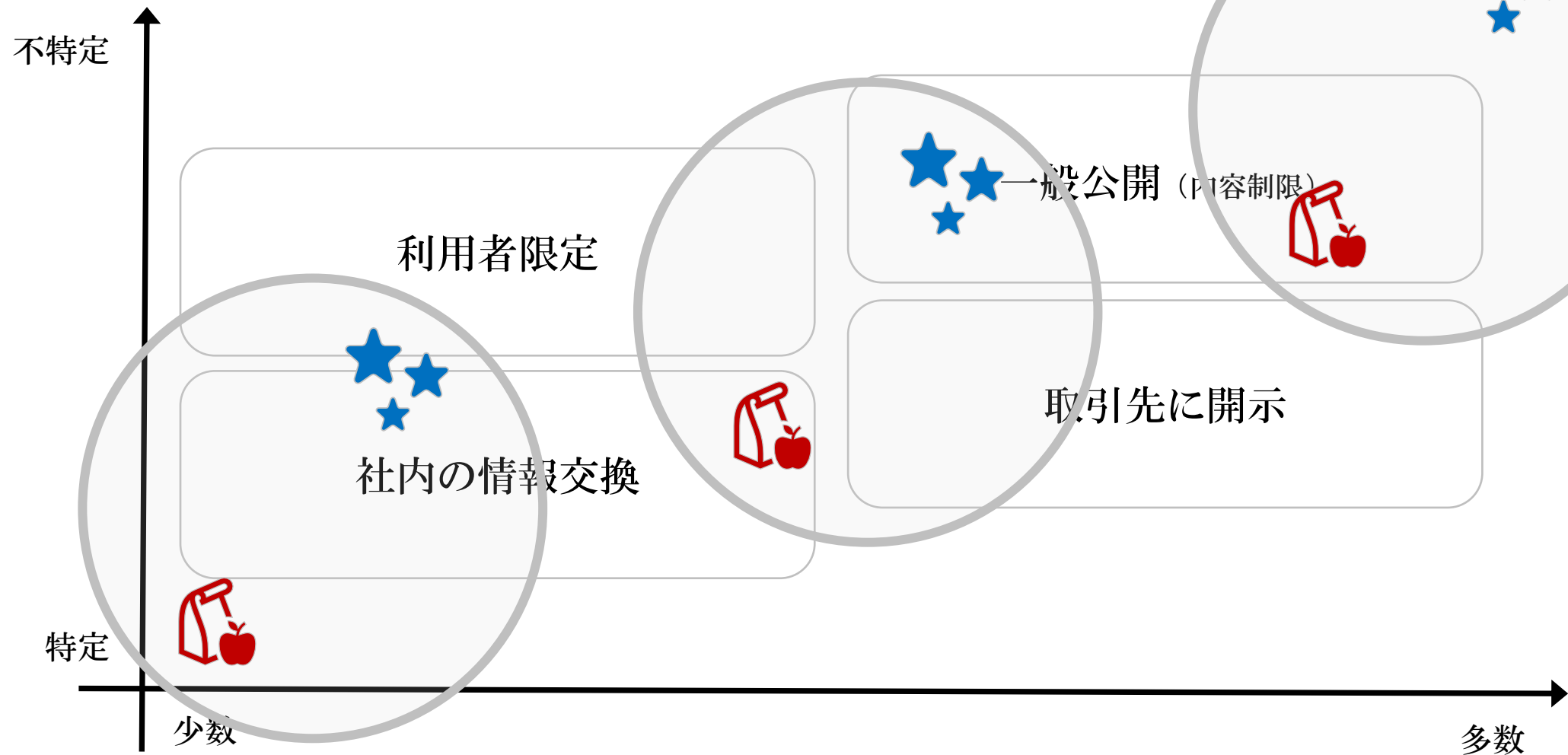


考察：出品登録・利用リクエスト

- **利点**：慣れれば簡単
- **難点**：利用者がまだ少ない

「鍵のかかった部屋」をいかに持ち運ぶか

- フルオープンと制限公開の境界線を「ずらす、動かす、増やす」と考えられないか



論点まとめ

1. 不特定多数と特定少数の「間」に多様なデータ製品の形態
 - 暗黙知や社会的慣習のちがいによって、制約もそれぞれ
2. 制限公開が選ばれる理由は複合的で一様でない
 - 法的課題、技術的問題、ブランド毀損・ビジネス機会損失リスクなど
3. 従来手法も通用するが、ルール&システムで省力したい
 - 「身内になる」と「一般開放（制限付き）」の2択以外の道が開ける

「鍵のかかった部屋」をいかに持ち運ぶか
