

2020年「NDLデジタルライブラリーカフェ」

ウェブアーカイブの 市場価値とビジネス利用をめぐって

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12月10日

株式会社日本データ取引所



1

会社紹介

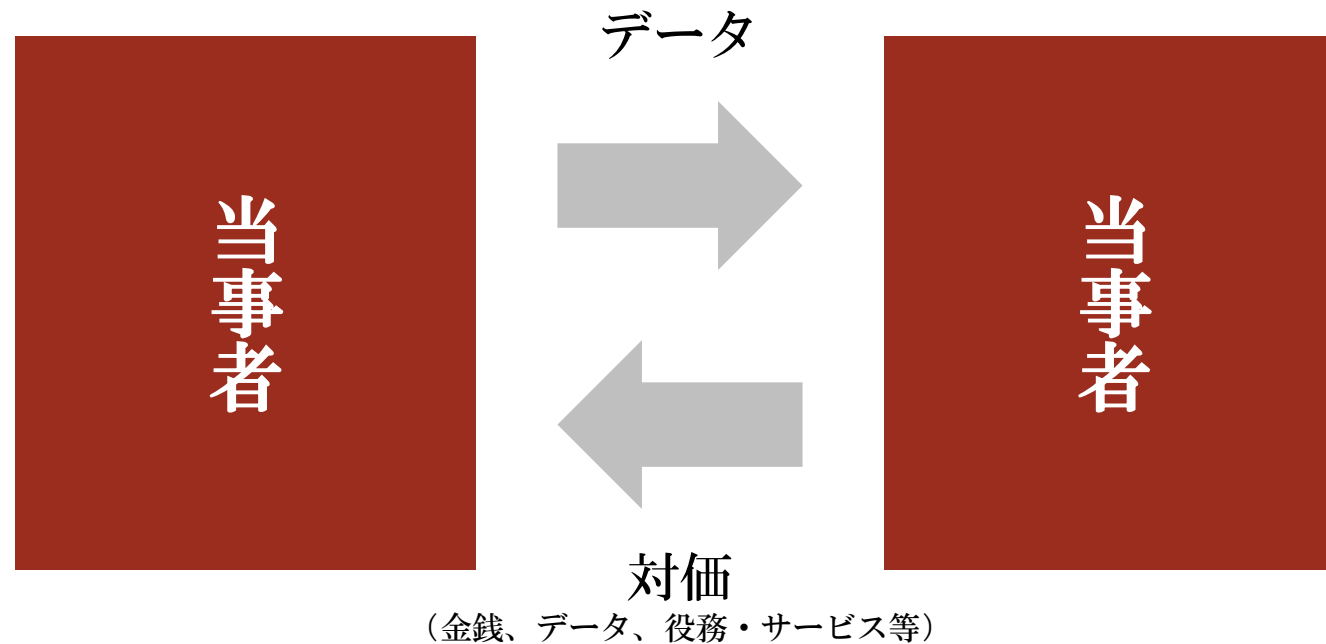
JDEX

買える！
データの図鑑



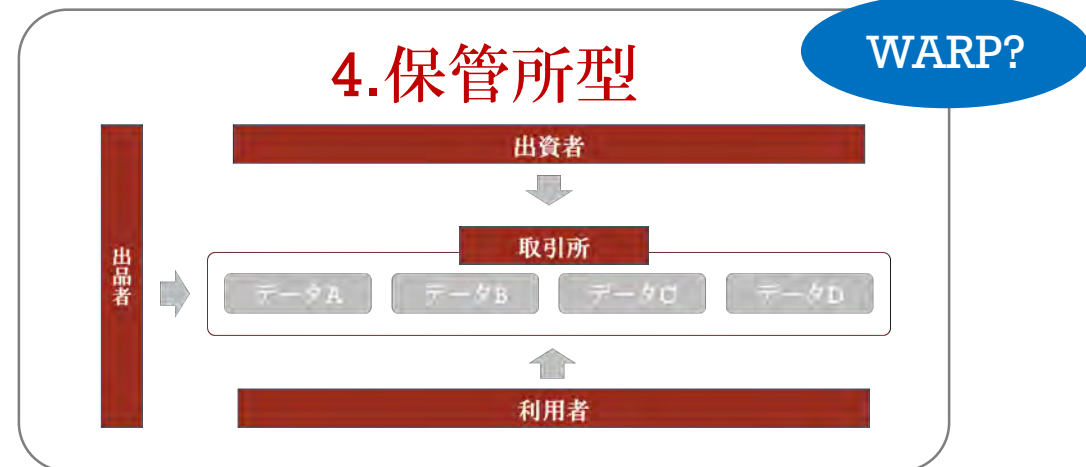
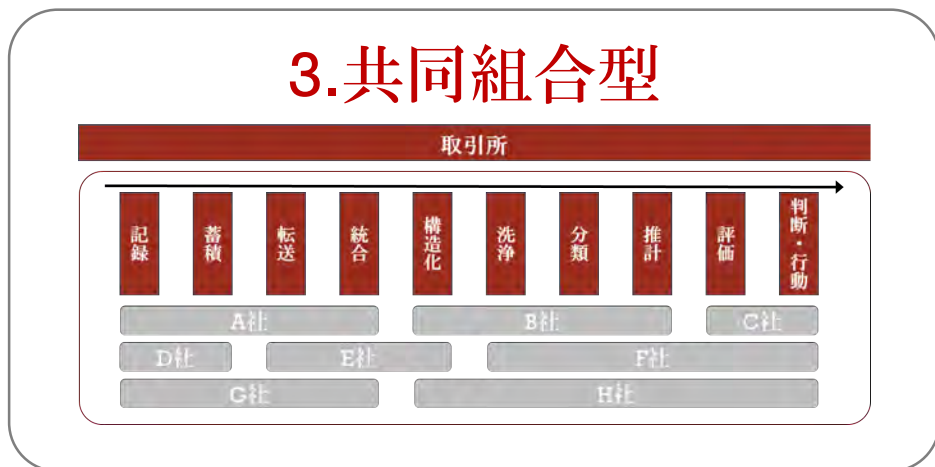
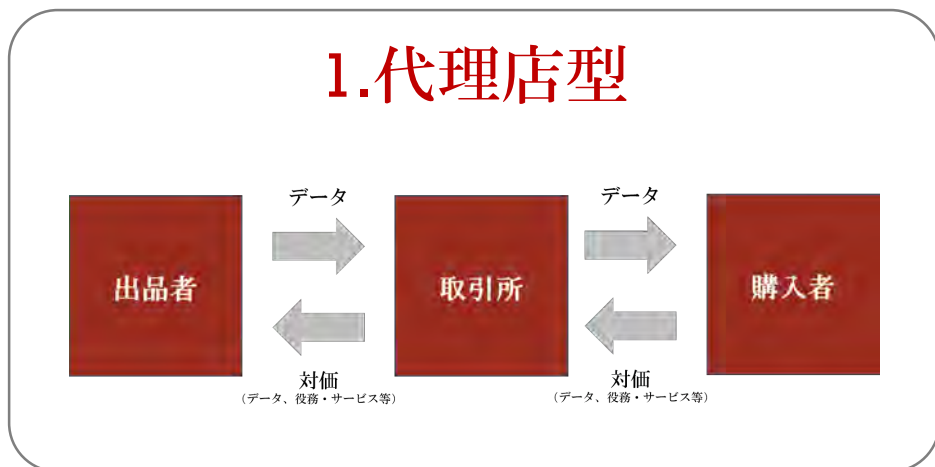
データ取引とは

- 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。
- 信頼できるデータを手入手する方法のひとつで、対価が金銭とは限りません。
- 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。



データ取引所とは何か

- 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。



本日の話題

1. 市場価値
2. データの流通性
3. ビジネス利用

本日の話題

1. 市場価値
2. データの流通性
3. ビジネス利用

データ流通市場とは

- データ製品は、他の財・サービスと同じように、多くの工程を経て作られます。
- 狭義の売買市場に限らず、一連の過程をデータ流通市場と総称します。

原料生産

製品加工

消費

記録

蓄積

転送

統合

構造化

洗浄

分類

推計

評価

判断・行動

市場価値とは（辞書的な定義）

市場

1. 商品としての財貨やサービスが交換され、売買される場についての抽象的な概念。

価値

- 1.物がもっている、何らかの目的実現に役立つ性質や程度。値打ち。有用性。
- 2.商品の価格の背後にあって、それを規定しているもの。その本質のとらえ方によって客観価値説（労働価値説）と主観価値説（効用価値説）とが対立する。

三省堂「大辞林」第四版（2019）

「実現目的」は利益追求に限らない

- 市場価値といっても、その価値が利潤の創出に限らないことはよく知られています。
- さまざまな種類の有用性がありうるので、市場価値の評価法も複数ありえます。

- 商業利用
- 報道活用
- 学術貢献
- 教育・福祉の充実
- 公共の便益
- 文化芸術の支援...etc.

「有用性」は「部位」ごとに考える

- ウェブアーカイブデータの有用性は、そのデータを構成する「部位」ごとに考えられます。
- 活用方法に応じた要素技術があり、多くの企業・個人が多彩な有償サービスを展開しています。

構成要素	テキスト	動画像	文書構造	URL	日付	運営者 (Whois)
活用例	目視による閲覧		文書抽出	メディア リスト	年表	辞書 ・名簿
	自然言語 コーパス	画像認識 物体検知	ページ 品質評価	ネット ワーク 分析	時系列 分析	識別子 ・分類

価値算定の考え方

- ひとつの目安として貨幣換算による価値算定もよく行われます。
- 分野ごとにいくつかの手法が考案・実践されています。

市場取引される財

- 費用から：コストプラス法、原価法
- 相場から：競合基準法、事例比較法
- 期待から：収益還元法、需要価格法

市場取引されない／しづらい財

- 保護価値の試算：仮想評価法（CVM）
- 経済波及効果の測定：
 - （参考例）ユネスコ文化統計局「文化産業の経済的貢献度の測定：現在の方法論的アプローチのレビューと評価」

WARPの場合は？（著者私見）

- ウェブアーカイブの市場価値を考えると、最終成果物のほかにも判断対象はあります。

- 選定の価値

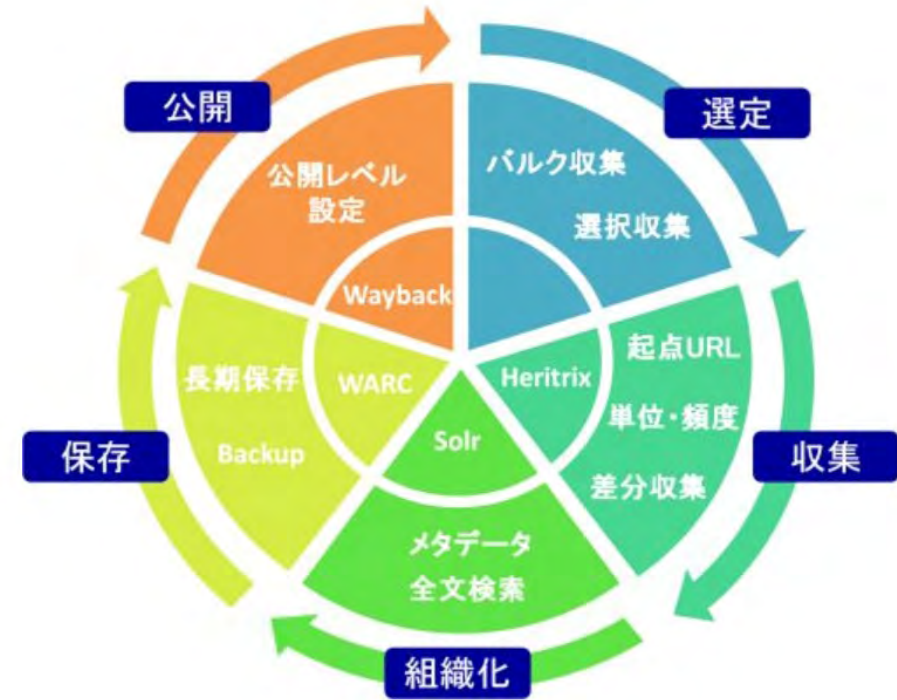
- 原材料（Webページ≒著作物）

- 収集・組織化の価値

- Webクロール
 - データクレンジング
 - メタデータマネジメント

- 保存・公開の価値

- （狭義の）アーカイブの価値
 - 二次利用による付加価値



ウェブアーカイブのライフサイクル

画像出典：国立国会図書館「ウェブアーカイブのしくみ」
<https://warp.da.ndl.go.jp/contents/recommend/mechanism/mechanism02.html>

本日の話題

1. 市場価値
2. データの流通性
3. ビジネス利用

流通性（流動性）とは

流動性

- ある資産を損失なく貨幣に代える容易さの度合。

出所：三省堂「大辞林」第四版（2019）

流通性

- 〔不動産用語〕土地・建物の売りやすさの度合。

出所：不動産流通推進センター「価格査定マニュアル」紹介ページをもとに著者作成

☞ データの流通性を高めるとは？

データの流通性をどう考えるか

- 既存の財・サービスに関する知識体系が応用できるのではないかと考えています。



参考になりうるのは、たとえばマーケティング戦略、データ品質評価指標、不動産国際評価基準、図書館パフォーマンス指標、SEO...etc.

ウェブアーカイブデータの流通性は？

- ウェブアーカイブ自体の流通性と合わせて、データ製品としてのそれも考慮できそうです。

知る	<ul style="list-style-type: none"> ウェブアーカイブといえは？（純粹想起） 〇〇なら、ウェブアーカイブ（ジョブの連想）
訪ねる	<ul style="list-style-type: none"> 検索ランクは？（いわゆるSEO） サイト内行動による流入は？
調べる	<ul style="list-style-type: none"> 初心者向けのガイド、ヘルプは？ 専門家・関係筋の利用例、評判は？
入手する	<ul style="list-style-type: none"> 提供サービスは充実している？ （検索サービス、データセット、オーダーメイド集計...etc.）
使う	<ul style="list-style-type: none"> データ内容の品質は？ データ内容に影響を与える諸条件の品質は？

WARPの場合は？

- 使用適合性の観点では充実しており、他の特性についても発展段階にあるとお見受けします。

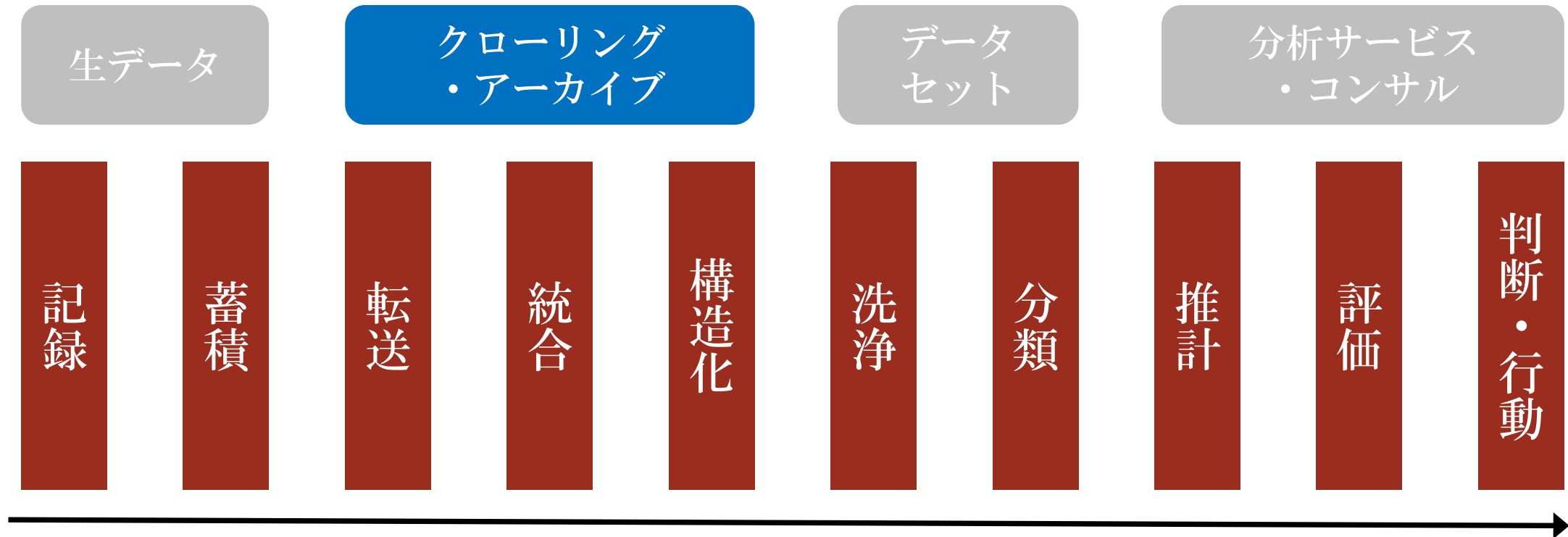
知る	<ul style="list-style-type: none"> • 国民一般の認知は？ • 図書館・アーカイブ業界、IT業界の認知は？
訪ねる	<ul style="list-style-type: none"> • 検索順位が高い、または重要なキーワードは？ • 各ページのPV数は想定に見合う？
調べる	<ul style="list-style-type: none"> • FAQ、おすすめコンテンツはすでに充実 • データセットの充実は今後、検討？
入手する	<ul style="list-style-type: none"> • 一般向けには検索サービスとして提供中 • 他の利用形態は今後、検討？
使う	<ul style="list-style-type: none"> • 大分類やクローリング結果のメタ情報はある • ファイル種別や可視化事例もある

本日の話題

1. 市場価値
2. データの流通性
3. ビジネス利用

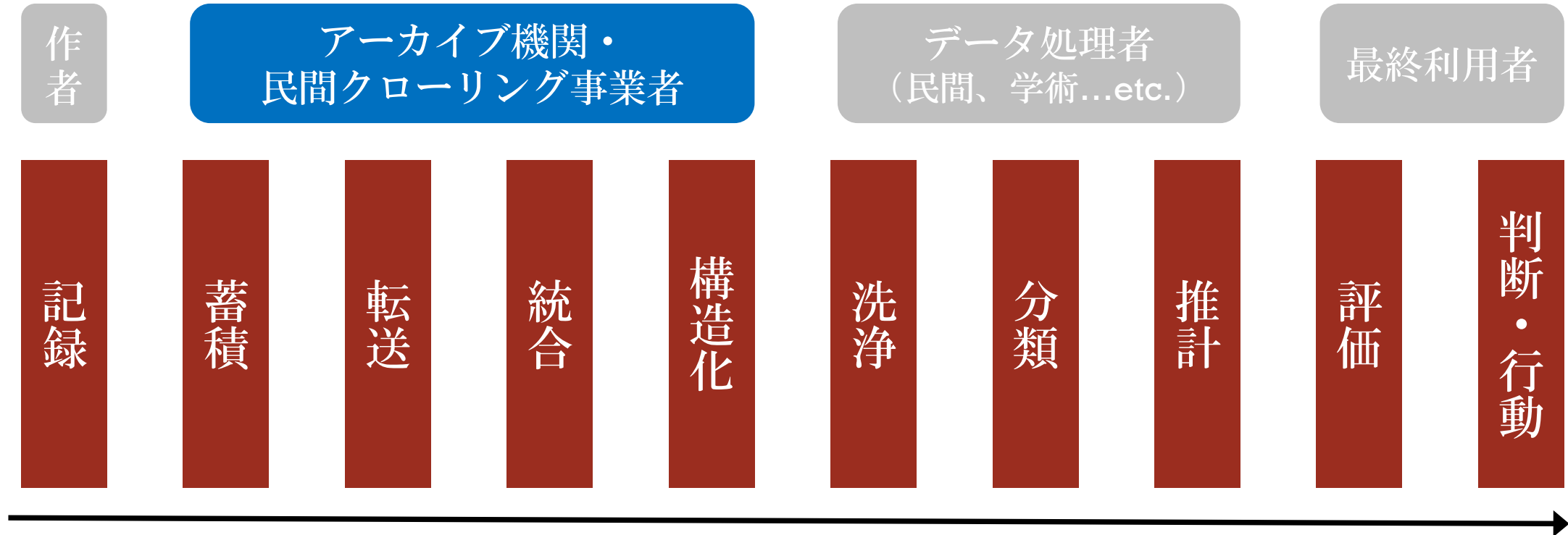
データ製品のさまざまな「形」

- データ流通の過程では、さまざまなデータ製品が作られます。
- 生データのみが売買対象となることもあれば、分析サービス化しないと買い手がつかない場合も。



社会的な役割分担をどう考えるか

- アーカイブ機関が作成したウェブアーカイブを製品加工し、最終利用者に届ける流れを考えます。
- データ処理者との協力、他のクロール事業者との棲み分けは、今後の論点でしょうか。



公的アーカイブのビジネス利用に向けて

- 収益化に直接貢献するビジネス（商売）に限らず、複数のビジネス（業務）での利用が期待されます。
- アーカイブから派生し、データセット販売や分析サービス・コンサル提供する例も多々あります。

収集対象	クロールング ・アーカイブ主体	利用条件	利用目的
公的機関の サイト	公的アーカイブ	オープン データ	レファレンス サービス
学術機関の サイト	学術リポジトリ	研究利用	論文・レポート
民間企業の サイト	クロールング事業者 個人による蒐集	商用利用	分析サービス・ コンサルティング

ビジネス利用の想定シナリオ（例）

- 「関心」段階のもの、具体的な「質問」があるものなど、データ製品への期待は多岐にわたります。
- 個別のアーカイブ対象ページ自体だけでなく、時系列情報や目的別統計などの需要も想定されます。

関心

新型コロナウイルス感染症にまつわる報道・世論の動きを網羅的に知るための略年表を作成しています。武漢市の鮮魚市場関係者の感染が発覚した時点から、最直近までの期間を対象に、日別のできごと年表を作成したいと考えています。この作業に役立つデータを探しています。形式や内容は問いません。

質問

- 省庁・自治体によるCOVID-19関連ページの増加数を週別で知りたい。
- 文化芸術団体による祭典・フェスティバルの開催／中止率を日別で知りたい。

まとめ

1. ウェブアーカイブの**市場価値**は複数の手法で評価できる。
商用利用による最終製品の貨幣換算額のみに限られない。
2. データの**流通性**は複数の観点で考えられる。
考慮事項は想起容易性から使用適合性まで幅広い。
3. **ビジネス利用**の方法は「関心」「質問」ごとに考える。
社会的な役割分担をどう構築することが理想的か？

結びに代えて—思考実験の素材

- ウェブアーカイブデータを市場流通させるとしたら、どのような製品企画がありうるでしょうか？

- (仮称) 新型コロナウイルスに関する日本語ウェブアーカイブデータ
 - 概要：選択収集の対象とする公的機関が公表したウェブページのうち、「原因不明肺炎」「新型コロナウイルス」「COVID-19」を含むもののテキストデータであって、適法に二次利用可能なもの
 - 対象期間：2020/01/01-2021/3/30
 - 対象地域：日本（日本法が適用されるウェブ空間）
 - 対象言語：日本語
 - 収録内容：データセット、データ項目定義書
 - 収集者：国立国会図書館
 - 提供者：データ処理者A
 - 利用条件：第三者提供の禁止、出所の表示、利用報告…etc.
 - 免責・非保障：データは現状有姿で提供され、内容品質は非保証
 - 利用料金：XXX, XXX, XXX円



ご視聴ありがとうございました。