

---

OAI-PMHと

メタデータ・ハーベスティング

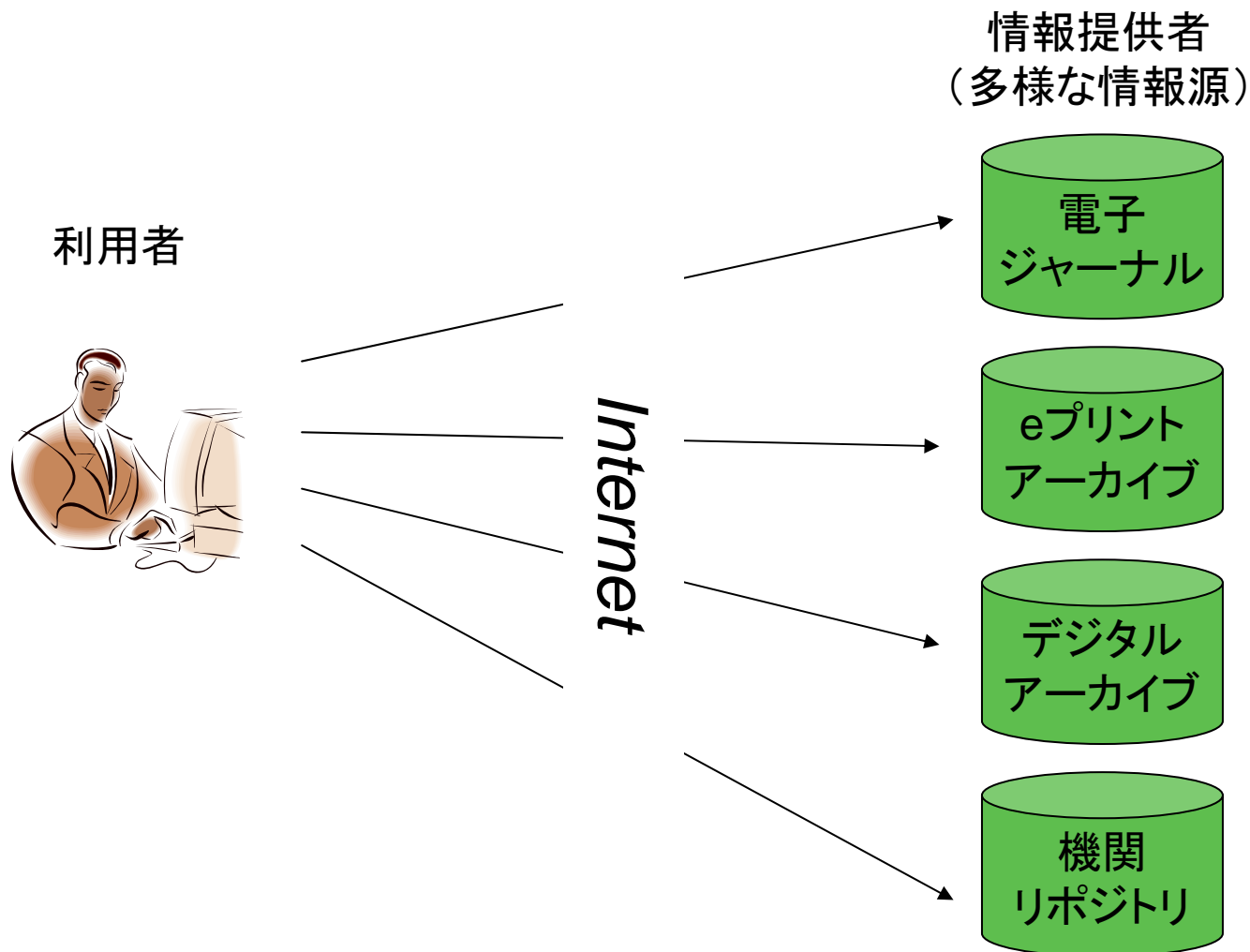
---

国立情報学研究所

開発・事業部コンテンツ課長

尾城 孝一 (ojiro@nii.ac.jp)

# 現在の電子情報環境における情報探索



# 問題の所在

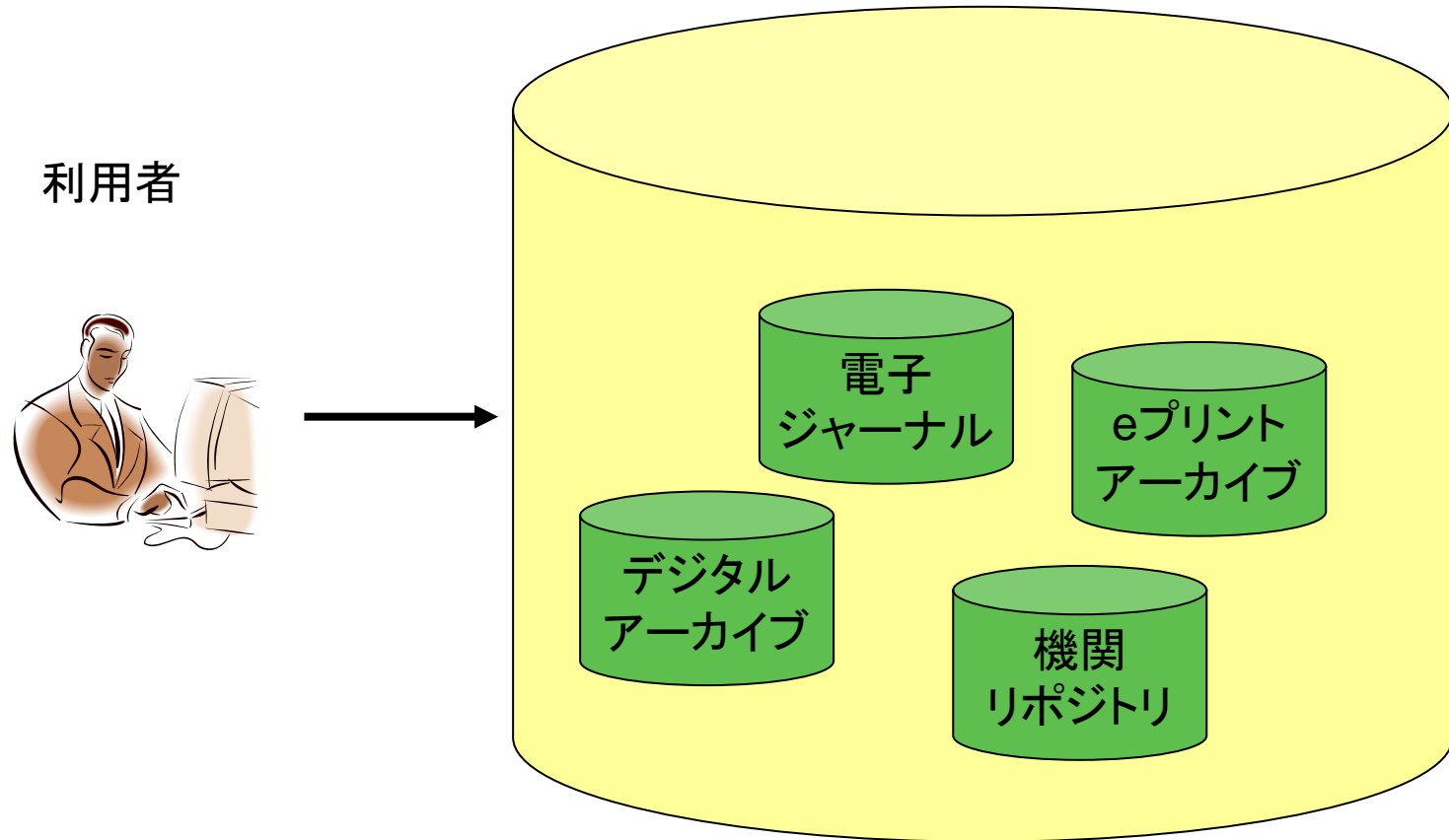
## ■ 利用者にとって

- 何がどこに？
- どのようにアクセスするのか？
- まとめて検索できない
  - 情報探索のコスト高

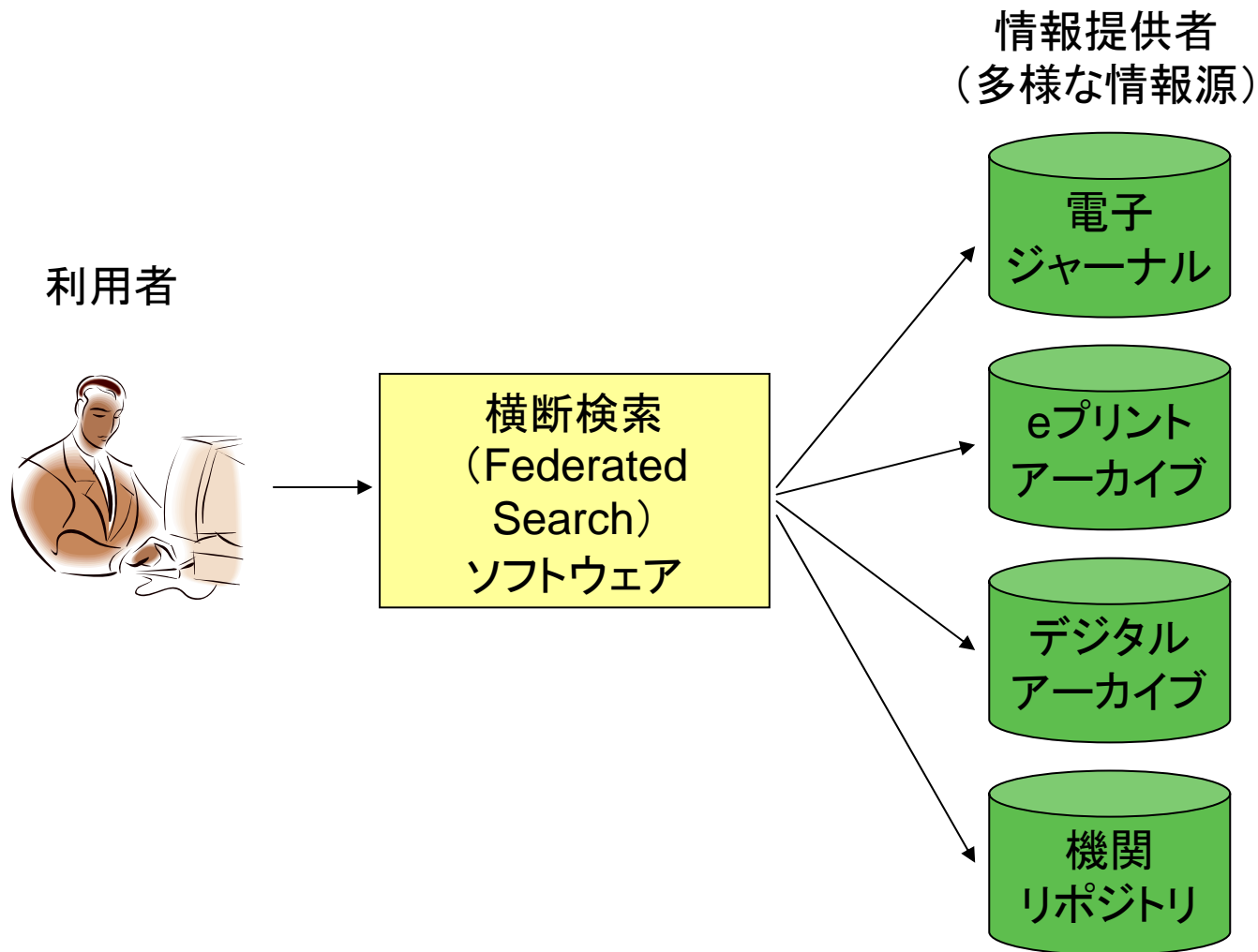
## ■ 情報提供者にとって

- 直接来訪者にしかサービス提供できない
- 膨大の情報の中に埋没
  - 提供コンテンツのインパクトが上がらない

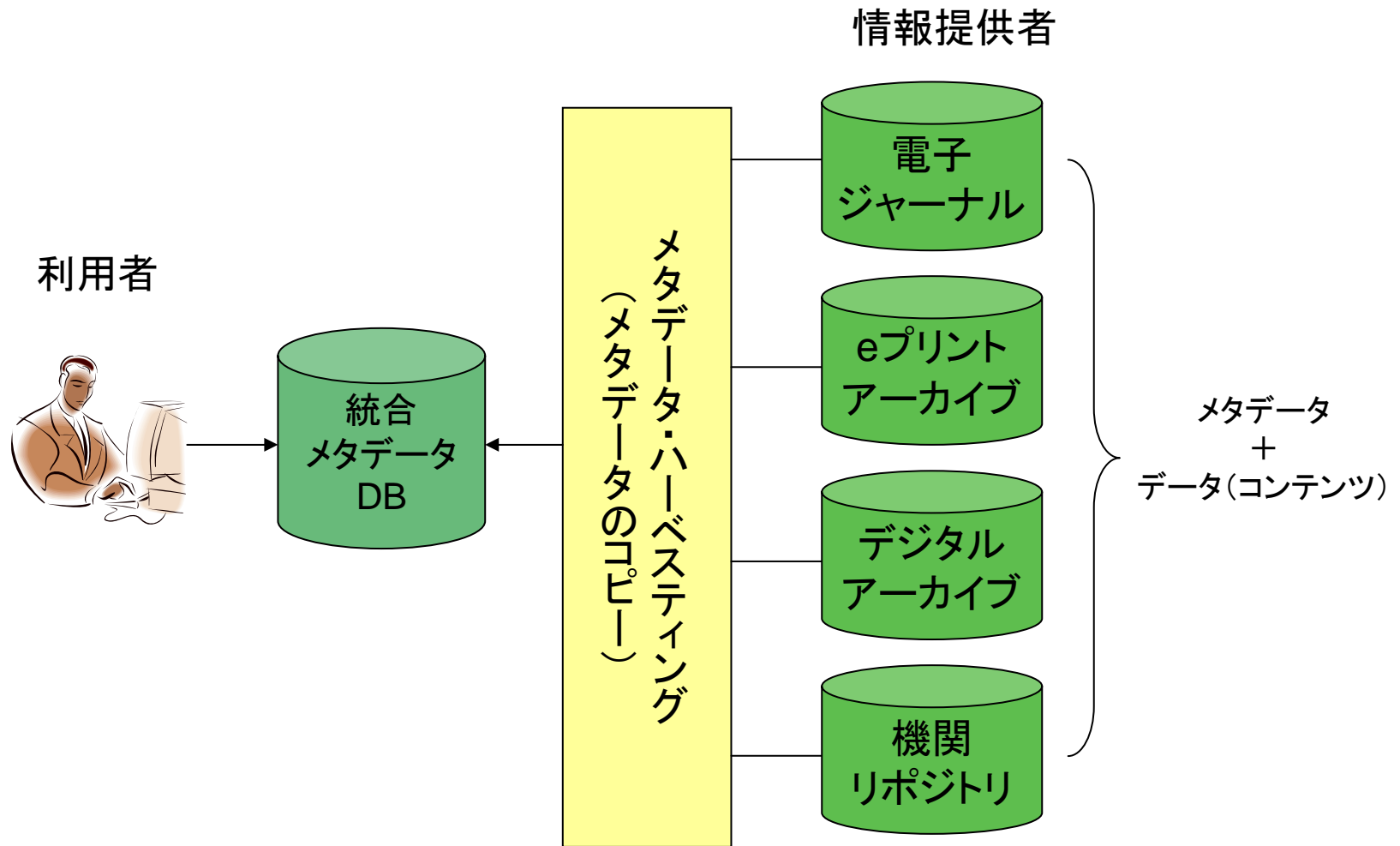
# 解決策1: 集中型ソリューション



# 解決策2: 横断検索 (Federated Search)



# 解決策3: メタデータ・ハーベスティング



# OAIとは



- OAI (Open Archives Initiative)
  - <http://www.openarchives.org/>
  - 「メタデータ・ハーベスティング」を通じて多様な電子情報サーバ間の相互運用を促進することを目的とした国際的な運動」
- Open ?
  - 「無償」ではなく「開かれたシステム」
- Archive ?
  - 厳密な意味での「アーカイブ」(保存庫)ではなく、広い意味での電子情報を蓄積するためのサーバ

# OAIの歴史

- 1999年10月（サンタフェ会議）
  - eプリント・サーバ（電子論文サーバ）の代表者の集まり
  - メタデータ・ハーベスティングに基づくフレームワーク（サンタフェ協定）
  - OAIの結成
- 2001年1月
  - メタデータ・ハーベスティング・プロトコル（Open Archives Initiative Protocol for Metadata Harvesting: OAI-PMH）Ver.1.0が制定
  - メタデータを刈り取る側と刈り取られる側の間の約束事（規約）
- 2002年6月
  - OAI-PMH Ver.2.0発表



# OAI-PMHフレームワークの構成者

- データプロバイダ
  - 電子情報サーバを維持し, OAI-PMHによりメタデータを開示する
- サービスプロバイダ
  - OAI-PMHを使用してデータプロバイダから収集したメタデータに基づき, 付加価値のあるサービスを提供する

# リポジトリとハーベスタ

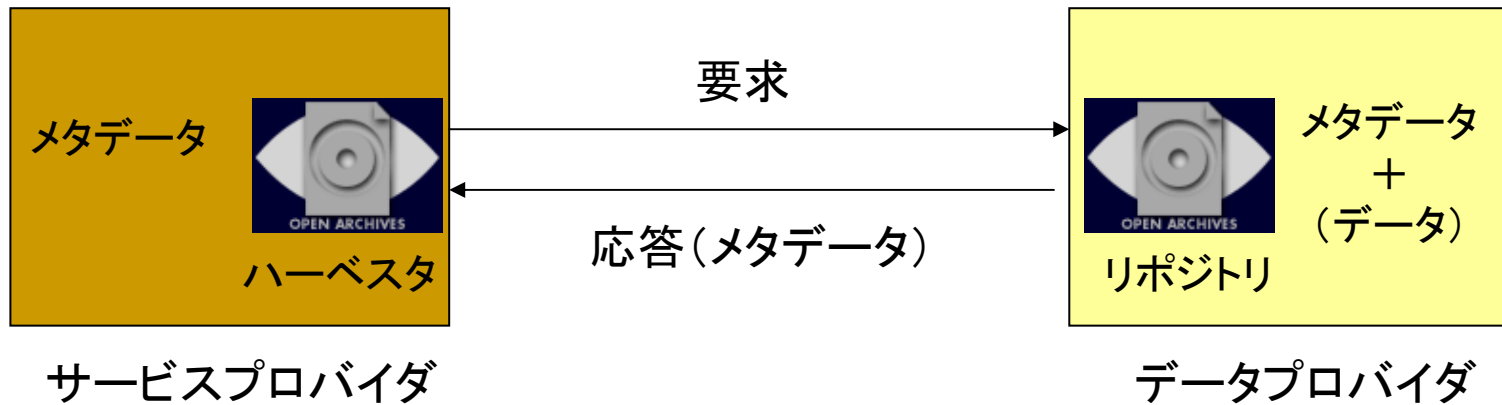
## ■ リポジトリ

- OAI-PMHの要求に対して応答することのできるネットワークアクセス可能なサーバ
- メタデータを開示するために、データプロバイダが管理する

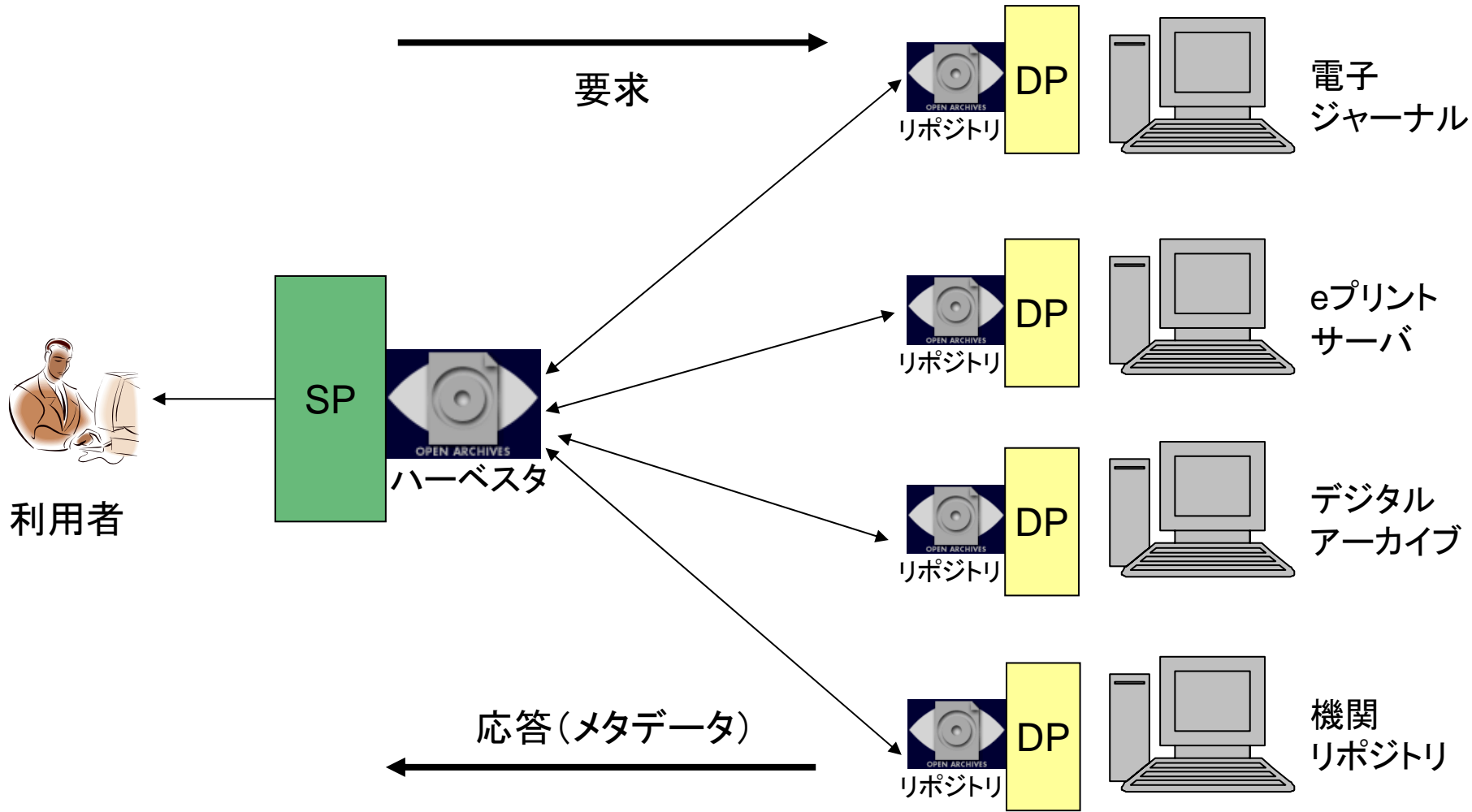
## ■ ハーベスタ

- OAI-PMH要求を発行するクライアント・アプリケーション
- サービスプロバイダがリポジトリからメタデータを収集する手段として使用する

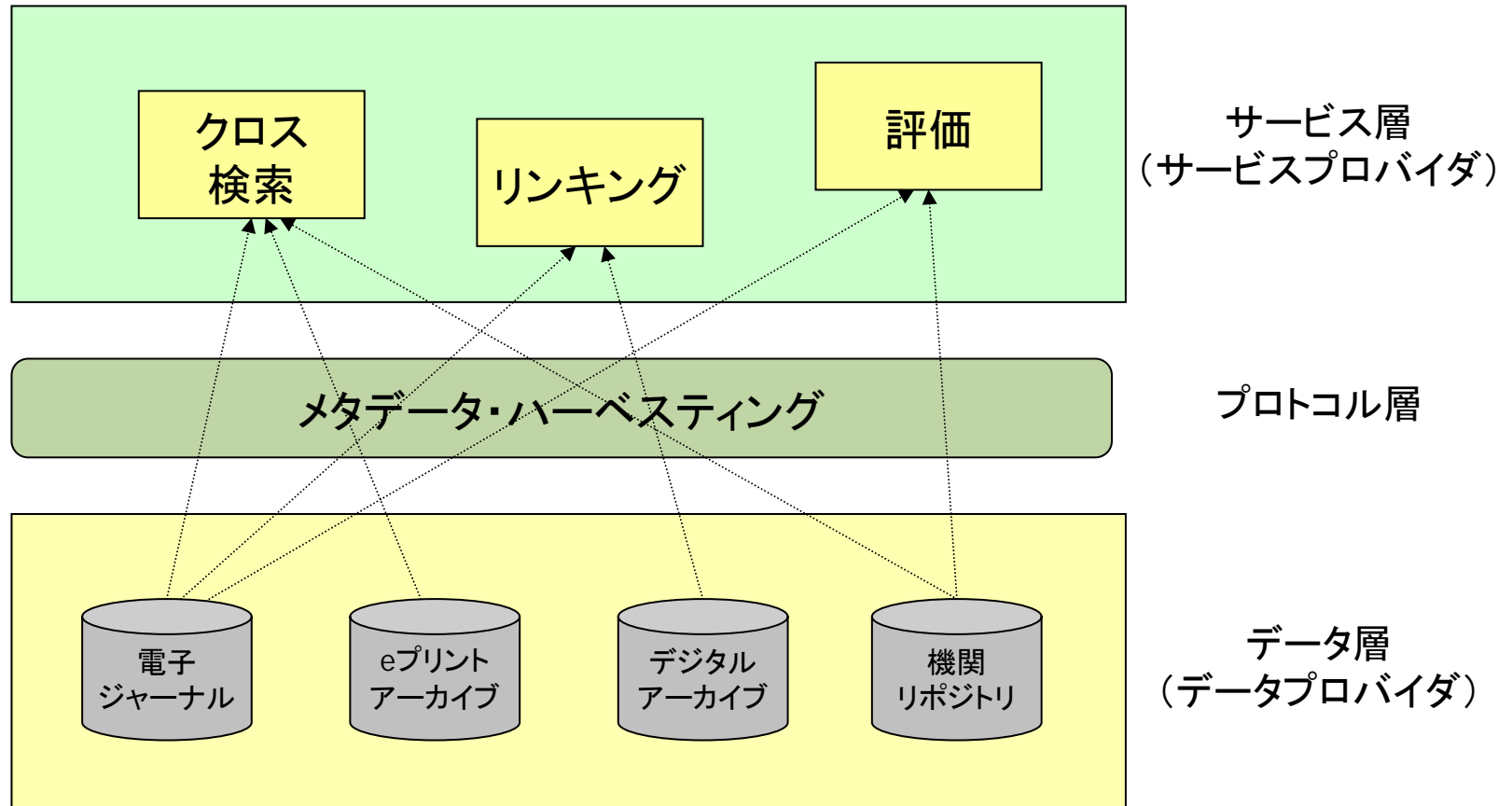
# データプロバイダとサービスプロバイダ(1)



# データプロバイダとサービスプロバイダ(2)



# OAIの3層モデル



# リソースーアイテムーレコード



「阿修羅像」に関する全ての  
利用できるメタデータ

ダブリンコア  
メタデータ

MARC  
メタデータ

JuNii  
メタデータ

リソース：  
メタデータが指し示す「物」  
デジタルでも非デジタルでもよい

アイテム：  
リソースに関する複数の  
メタデータを格納する概  
念的な入れ物  
一意の識別子を持つ

レコード：  
ある特定のフォーマットで  
表現されたメタデータ  
XMLでコード化されやりとり  
される  
Simple DCは必須

# セットと選択的収集

## ■ セット

- 選択的収集のために複数のアイテムをグルーピングする概念
- 必須ではない

## ■ 選択的収集

- セットと日付スタンプを指定して、選択的にメタデータを収集することが可能

# OAI-PMH要求

	要求(動詞)	機能
リポジトリに関する情報を得るための要求	Identify	リポジトリに関する情報を取得する
	ListMetadataFormats	リポジトリにおける利用可能なメタデータ・フォーマットの一覧を取得する
	ListSets	リポジトリのセット構造を取得する
メタデータ収集のための要求	ListIdentifiers	リポジトリからレコード中のヘッダー情報のみを取得する
	ListRecords	リポジトリから条件に合致するレコードを全て取得する
	GetRecord	リポジトリから個々のレコードを取得する



# プロトコルの特徴

- シンプルで「敷居の低い」プロトコル
  - 機能の制限
- 既存の標準の流用→実装コストの低減化
  - HTTP要求
    - OAI-PMHの要求は、HTTPのGETもしくはPOSTを使って送信される
  - XML応答
    - OAI-PMHの要求に対する応答は、全てXMLでエンコードされ、文字コードにはUTF-8が使用される

# 使用例 (GetRecord)

http://mitizane.ll.chiba-u.jp/cgi-bin/oai/oai2.0?  
verb=GetRecord&  
metadataPrefix=oai\_dc&  
identifier=oai:mitizane.ll.chiba-u.jp:00020115

ベースURL

要求の種類

引数

リポジトリ (<http://mitizane.ll.chiba-u.jp/cgi-bin/oai/oai2.0>) から  
識別子 `oai:mitizane.ll.chiba-u.jp:00020115` のレコードを  
`oai_dc` のメタデータフォーマットで取得せよ

# データプロバイダ

- 現在(2006年1月22日), 376のリポジトリが OAI-PMH(ver.2.0)に準拠したデータプロバイダとして登録されている
  - <http://www.openarchives.org/Register/BrowseSites.pl>

# サービスプロバイダ

- 現在(2006年1月22日), 23のサービスがOAI準拠のサービスプロバイダとして登録されている
  - <http://www.openarchives.org/service/listproviders.html>

# OAI-PMHの適用事例

## ■ データプロバイダ

### □ 千葉大学学術成果リポジトリ(CURATOR)

- <http://mitizane.ll.chiba-u.jp/curator/index.html>

- 千葉大学内で生産された電子的な知的生産物(学術論文, 学位論文, プレプリント, 統計・実験データ, 教材, ソフトウェアなどの学術情報)を蓄積, 保存, 発信するための電子書庫

## ■ サービスプロバイダ

### □ OAIster(「オイスター」)

- <http://oaister.umdl.umich.edu/o/oaister/>

- ミシガン大学のプロジェクト。500以上の機関からハーベスト600万件を越えるメタデータ・レコードの検索サービスを提供。

# 学会(誌)にとってのOAI-PMH

- インパクトの向上
  - 掲載論文の流通性, 可視性, 認知度を高めたい  
→ データプロバイダ
- ポータル構築
  - 関連分野のポータルを作りたい  
→ サービスプロバイダ

# OAI-PMHに関する情報源

## ■ NIIのOAI-PMH関連ページ

- <http://www.nii.ac.jp/metadata/oai-pmh/>
- OAI-PMHに関する技術情報, 関連文献等

## ■ NII軽井沢情報セミナー

- <http://www.nii.ac.jp/hrd/HTML/Joho-karuizawa/>
- OAI-PMHを活用した学術ポータル設計・試験構築