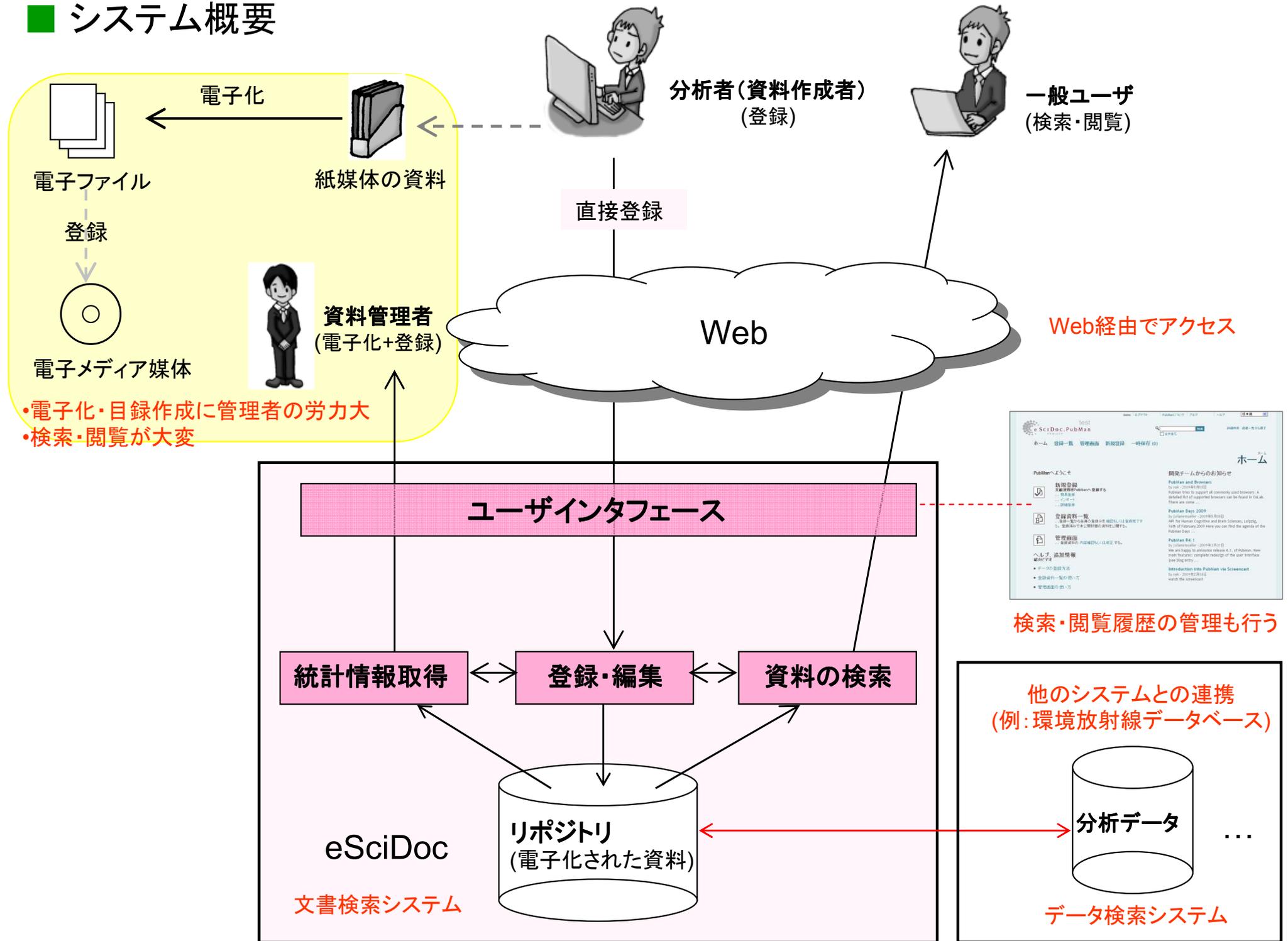


資料の電子化に関するご提案

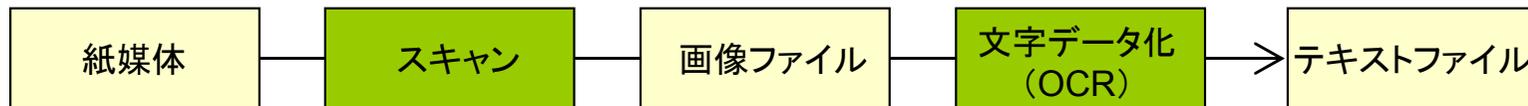
Advanced Algorithm and Systems

■ システム概要



■ 電子化する際に考慮すべき問題

1. 電子化にかかる費用



● 見積り例 (業者)

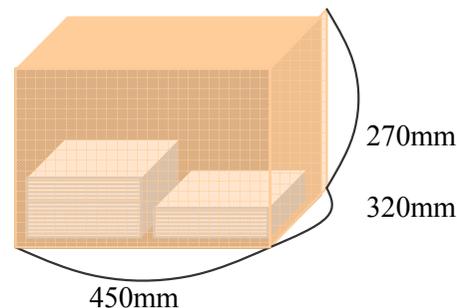
- ⇒ スキャニング (300dpi) : 1 ページあたり ¥7.9
¥7.9 × 5000 ページ = **¥39,500**
- ⇒ スキャニング + OCR 処理 : 1 ページあたり ¥30
¥30 × 5,000 ページ = **¥150,000**

● 見積り例 (自前)

- ⇒ スキャナ
価格
¥200,000 ~ ¥800,000 程度
- 処理時間
未定

⇒ OCR ソフト

- 価格
¥2000 ~ ¥20,000 (個人用)
¥200,000 ~ 350,000 (業務用)
- 処理時間 (個人用ソフト使用)
0.9分 × 5000 ページ = **75時間 (約9 ~ 10人日)**



A4サイズ用紙5000ページ (約20kg) 収納



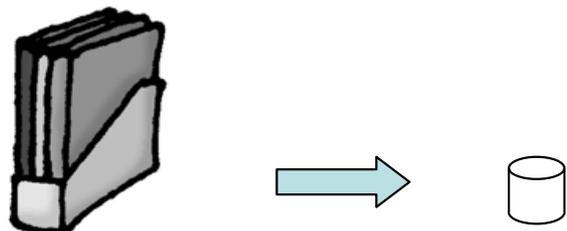
850MB: 画像ファイル (スキャン後)
75MB: テキストファイル (OCR後)
認識精度 95%

2. 効率的に電子資料を扱う方法

- 閲覧の効率化。→ **Web経由 (いつでもどこからでも)**
- 検索の効率化。→ **資料内容を正確に反映した目録 (メタデータ) の作成**

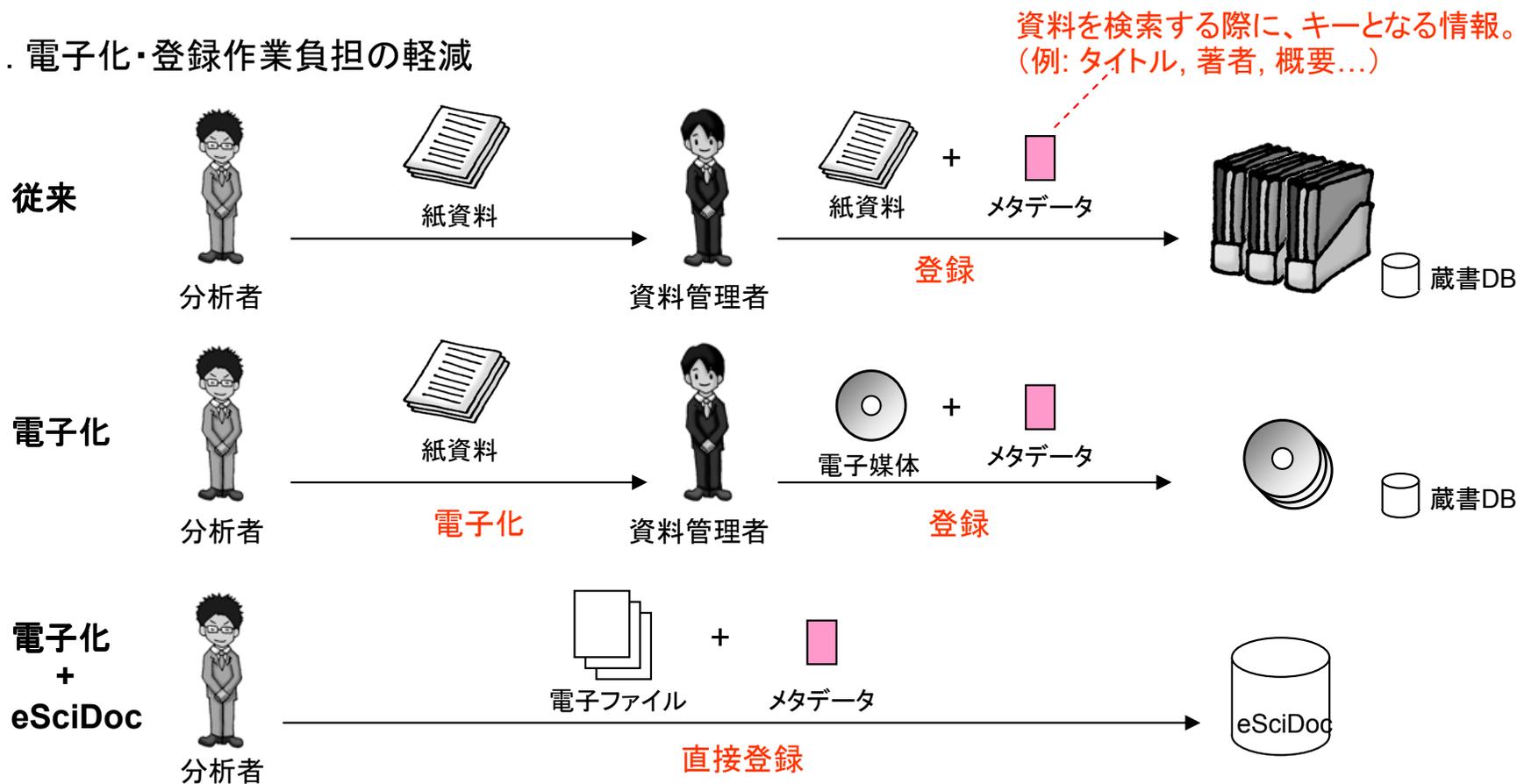
■ eSciDoc導入による資料管理者のメリット1

1. 省スペース化



2. 分析者(資料作成者)に直接登録を行ってもらうことが可能になる

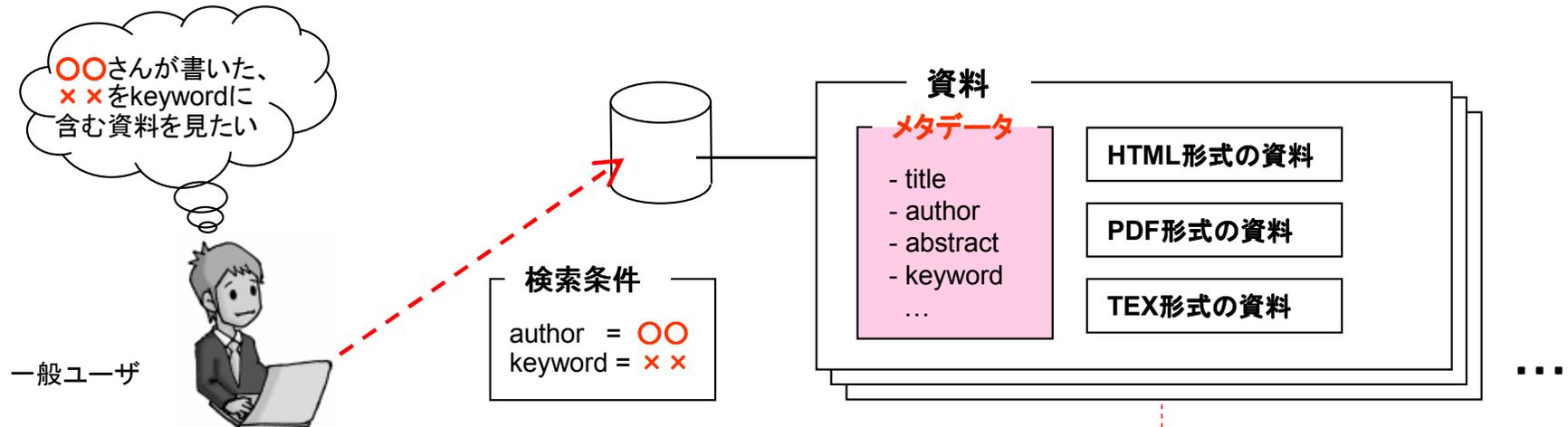
2-1. 電子化・登録作業負担の軽減



■ eSciDoc導入による資料管理者のメリット2

2-2. メタデータ精度の向上

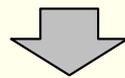
メタデータと検索



eSciDocは他のリポジトリシステムと比べて、
メタデータの定義が柔軟に行える

例えば ...

- ・ 資料種別 (報告書, 分析データ, ...)
- ・ どの部署が発行した資料なのか



使用状況に合わせたメタデータを
定義することが可能



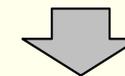
分析者(資料作成者)



資料管理者



分析者の方が資料内容に関する知識がある



分析者がメタデータを作成することにより、
メタデータの精度が上がる

■ eSciDoc導入による資料管理者のメリット3

3. 初期投資が小さくて済む

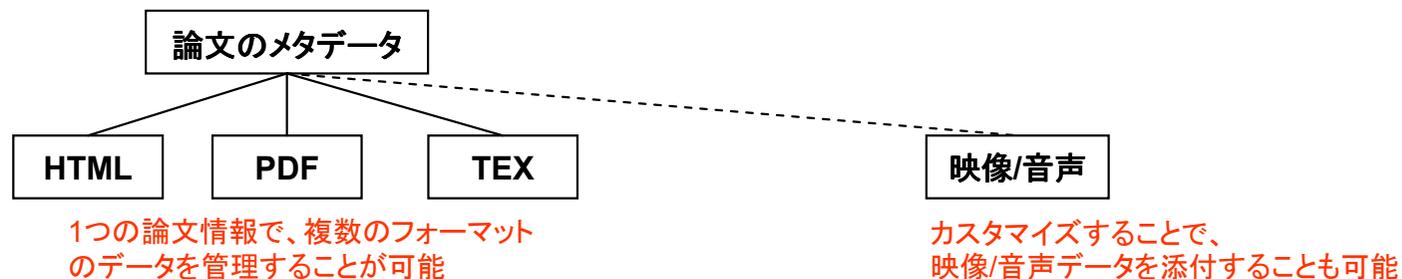
- ・eSciDocは、無料で使用できる
- ・ソースコードが公開されているので、カスタマイズすることが可能
(公開されているシステムの流用は、開発工数の削減にもつながる)

4. 既に導入実績がある

- ・物質・材料研究機構 や Max Planck Society で実際に使用されている

5. 様々な形式のファイルをまとめて管理

- ・音声や映像といったものも、文書と一緒に扱うことができる



6. 登録した資料へのフィードバック

- ・資料へのアクセス数などの情報が取得できる
⇒ 公開した資料(分析データ)がどの程度使用されたかを知る目安になる

■ eSciDoc導入による分析者(登録する人)のメリット

1. 目的に応じたSolutionがある

eSciDocには複数のSolution(ユーザインタフェース+拡張機能)があるので、目的に応じてそれらを使い分けることが可能

■ PubMan

- ・ 分析データや文献を、検索・登録するSolution
- ・ リポジトリへの分析データ・文献の登録が可能 (メタデータの登録・編集も可)
- ・ メタデータを使用した、柔軟なデータ検索が可能



■ ViRR

- ・ 電子化(スキャン)した資料を、ドキュメントの形にまとめるSolution
- ・ 電子化した資料に、keyword等のメタデータや注釈を付与することが可能



■ FACES

- ・ 画像データ共有のためのSolution
- ・ 顔写真の人物の年齢・性別・感情等の情報使用した、ファセット型ナビゲーション
⇒ 文書形式でない分析データを、コレクション管理するのに使用できる



■ eSciDoc導入によるユーザ(検索する人)のメリット

1. 資料を見るために移動しなくてよい

- ・ネットワークを介して、自分のパソコンで資料を見ることが可能になる

2. Myページ管理 (検索・閲覧履歴)

- ・履歴が残るので、資料の検索がしやすくなる。

3. メタデータを利用した柔軟な検索

■ eSciDocについて

Solution

PubMan (文献情報)

Web インタフェース

- ・ 登録
- ・ タイトル・本文検索
- ・ 統計情報
- ・ 文献情報エクスポート
etc...

ViRR (スキャン資料)

- ・ 目次/章/節構造に応じたナビ
- ・ 複数の専門家による
注釈の付与
etc...

FACES (画像)

- ・ 年齢・感情・性別による検索
- ・ メタデータ属性
- ・ ズーミング
etc...

共通機能の呼び出し

ネットワーク経由での呼び出し
に対応しているので、Solutionと
共通サービスを別マシンで動作
させることが可能

共通 API

eSciDoc (基盤共通サービス)

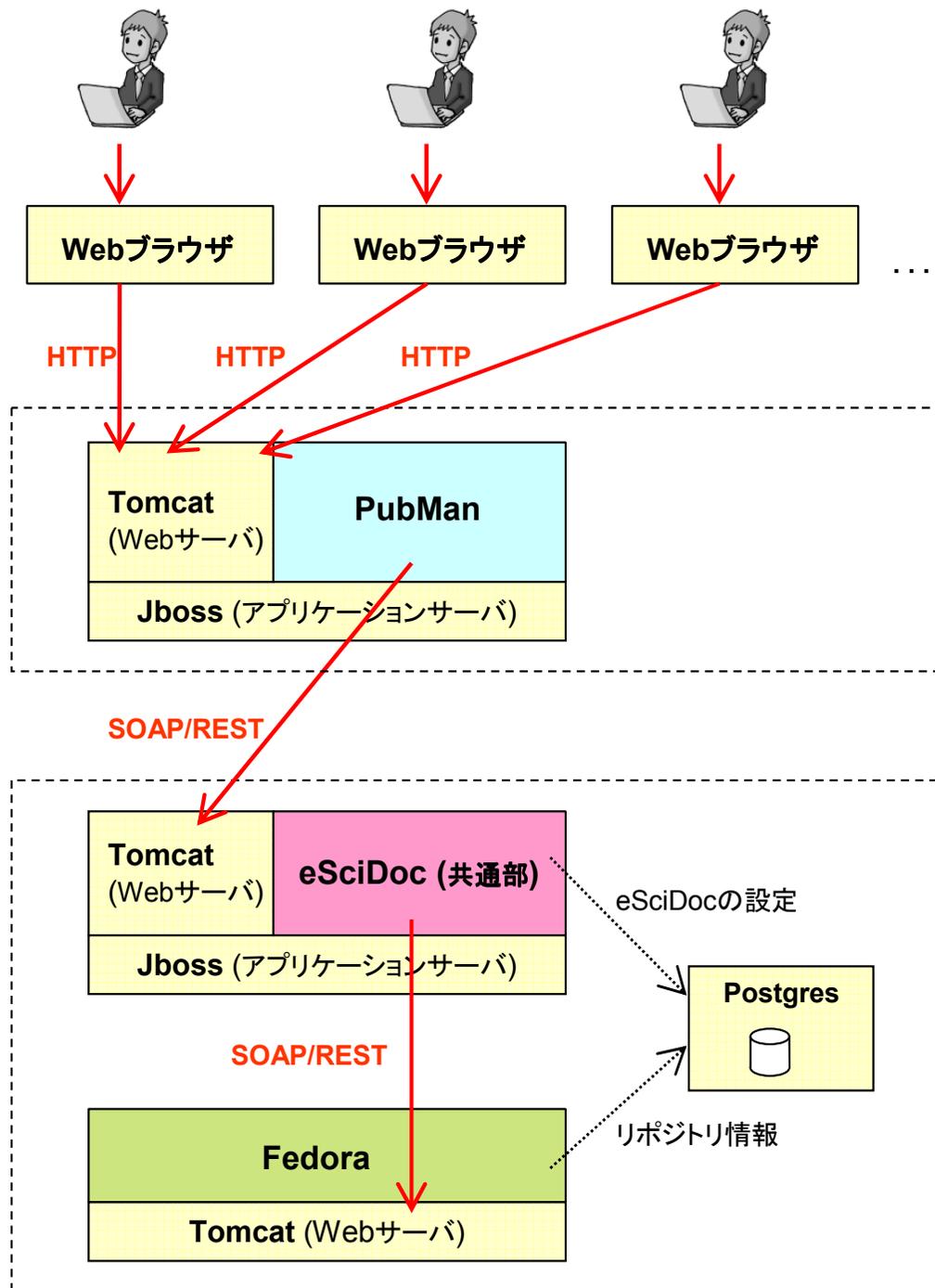
- ・ ログイン認証
- ・ 検索 (SRU/SRW)
- ・ 利用者統計情報
- ・ オブジェクト・所属管理
- ・ アクセス制御
- ・ 入出力
- ・ 画像Viewer環境

Fedora

リポジトリ管理

リポジトリ情報

■ eSciDocの動作イメージ



■ 見積もり(eSciDocカスタマイズ)

■ 検索・登録機能、Myページ、統計情報(Pubman流用) 【3人月】

- ⇒ 画面デザイン
- ⇒ 検索・登録フォーマット
- ⇒ 統計情報の種類
- ⇒ メタデータのフォーマット

■ 編集機能、図面・手書き資料等の電子書籍化、目次などによる構造化(ViRR流用) 【3人月】

- ⇒ 目次作成
- ⇒ 画面デザイン

■ 図面や写真など数値データ以外の検索(FACES流用) 【3人月】

- ⇒ 画像View環境
- ⇒ 検索フォーマット
- ⇒ 画面デザイン

■ ファイル検索⇔データ検索の連携(新規作成) 【1.5人月】

■ 日本語化(物材研と連携?) 【1.5人月】

■ 見積もり (システム・DB関連・メンテナンス)

● Escidocのシステムセットアップ作業

⇒インストール作業、各種設定 (5人日)

● DBバックアップ

⇒月次、年次のバックアップバッチの作成(ダンプファイルの自動出力機能)

⇒バックアップファイルの外部メディア出力、保持(DVD等へ書き出して外部保存)

● メンテナンス作業

⇒必要に応じたDBチューニング作業

⇒記憶容量の拡張

● トラブル対応 (ハードウェア・ソフトウェア)

⇒バックアップファイルからのデータ復旧等

⇒ハードディスクの換装等

● サーバダウン対策

⇒無停電電源装置(UPS)の導入(10万円程度)

⇒ Escidocのシステム再セットアップ等

● サーバ価格

