

機関リポジトリ(IR) システム概論

平成21年度学術ポータル担当者研修
H21.8.6(於 名古屋大学)
広島工業大学 森保信吾

本日の内容と注意事項

- 内容
 - 構築に必要なOSやWebサーバ(WWWサーバ)、データベースまで含めたIRシステム概要
 - IRシステムの種類
 - 自力構築 or 業者委託
 - セキュリティとバックアップ
 - IRシステムの諸機能(比較ポイント)
 - 各IRシステムの簡単な紹介
- 注意事項
 - 正確さを犠牲にして、分かり易さ重視で説明を試みます(実際に分かり易い説明であるかどうかは分かりません)

IRシステムの機能・必要条件

- Webインターフェースを提供すること
 - コンテンツを要求に応じて利用者に届ける機能
- メタデータとファイルが管理できること
 - コンテンツの保存と整理
- OAI-PMHに対応していること
 - プロバイダ機能(刈り取ってもらうための機能)

関連用語

- OS
 - 各ソフトが共通に必要なとする標準的機能を担う
 - ファイル管理、キーボード入力、ネットワーク通信など
 - Linux、Windows Server 2008、Mac OS X...
- 仮想マシン (virtual machine)
 - 物理的な機械の振る舞いをソフトウェアで真似て作ったあたかも機械があるように見える環境
- Webサーバ
 - Webサービスの窓口担当
 - Apache、Microsoft IIS...
 - TomcatはWebサーバではないが、ややこしくなるので今日のところはWebサーバみたいなものだとすることにする
- データベース
 - 様々なデータを扱い易いように整理保存する仕組み
 - MySQL、PostgreSQL、Oracle...
- プログラム言語
 - IRシステムなど、ソフトを開発するためのプログラム文法やツール(以下「呪文」)
 - JAVA、Perl、PHP、(HTML)...

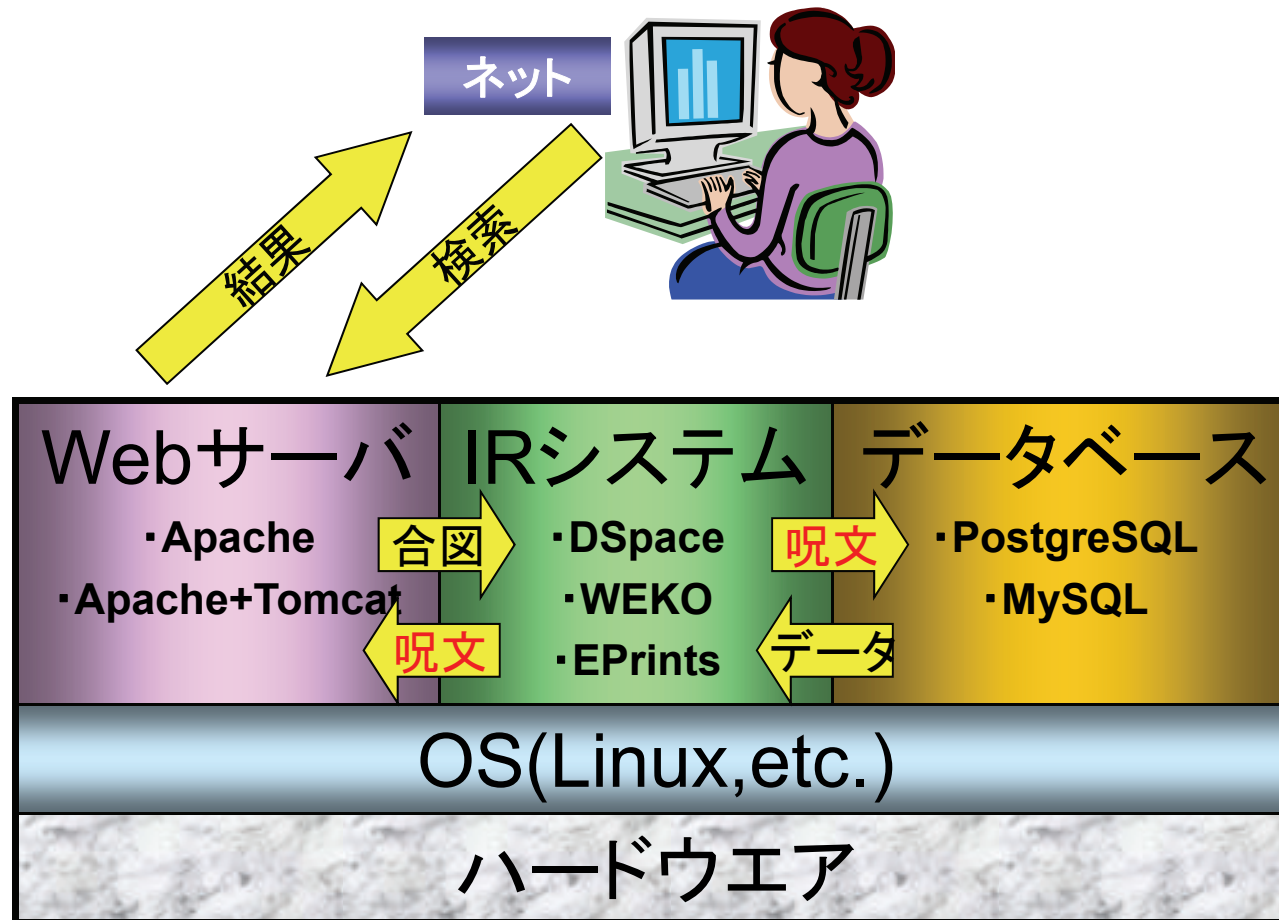
無理を承知で図書館に例えると

コンピュータ

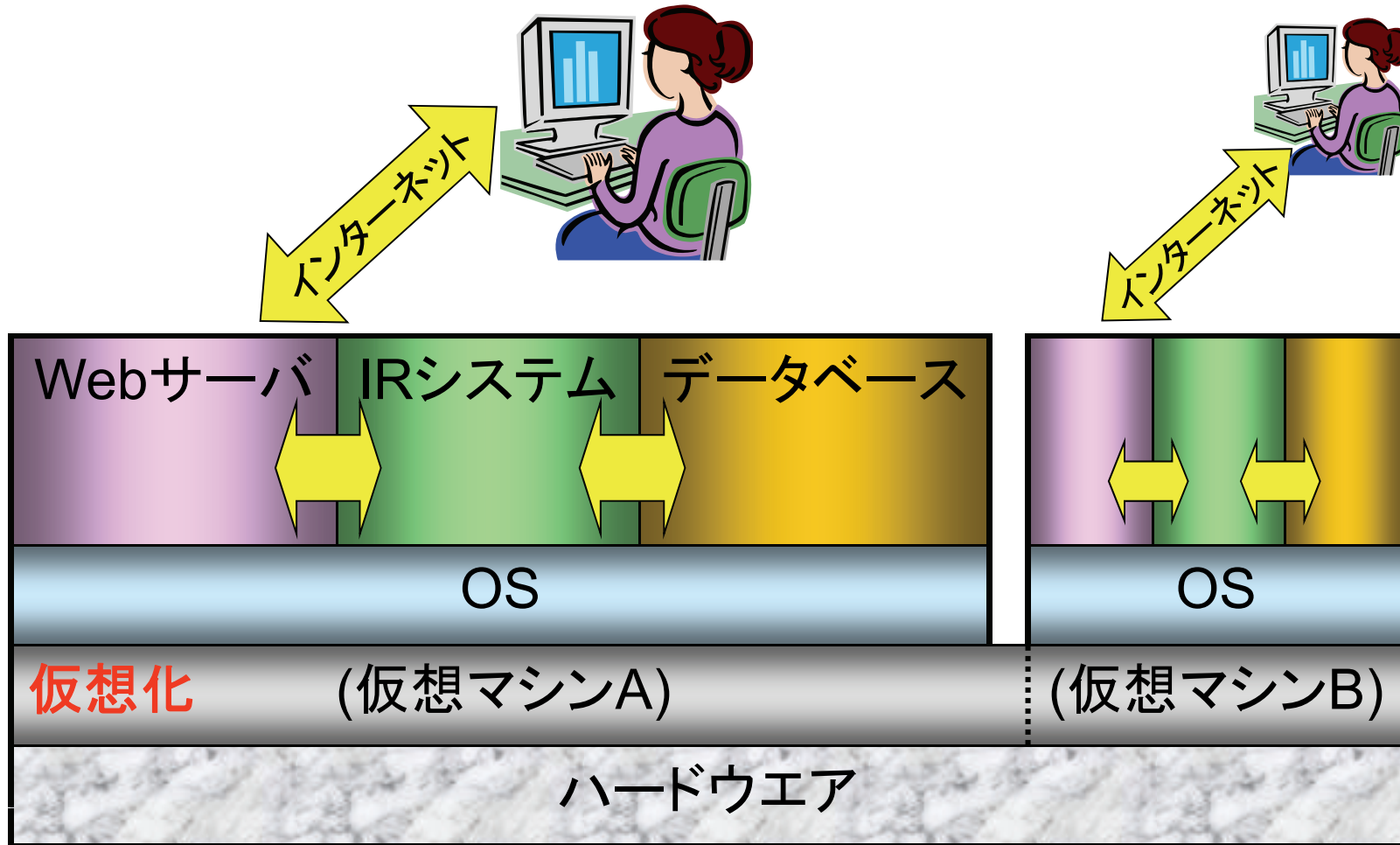
図書館

Webサーバ		データベース	参考受付	カード目録
O S	入力(キーボード等)、 提示(画面・プリンタ等)、 フォルダ・ファイル管理、 ネットワーク通信		受入係、 閲覧係、 書架整理、 図書館連携機能	
	ハードウエア (ハードディスク,メモリ,・・・)		建物 (書架,ブックトラック,・・・)	

IRシステムの概要-基本形



IRシステムの概要-仮想化



IRシステム(オープン・ソース)

名称	開発元 / 活動状況	国内シェア	開発言語 / DB
DSpace	MIT & HP / 活発	多 北大 他	JAVA PostgreSQL, Oracle
EPrints	サウサンプトン大学 / 活発	少 岡山大 他	Perl MySQL
WEKO	NII / 開発者が身近	?	PHP MySQL
XooNips	理化学研究所 / 割と活発	少 慶大 他	PHP MySQL
Fedora	コーネル大学ほか / 活発	?	JAVA PostgreSQL, MySQL 他

IRシステム(商用)

名称:	開発元	国内シェア	備考
E-Repository	CMS	少 千葉大 他	機関リポジトリのために開発されたシステム
NALIS-R	NTTデータ	中	DSpaceをベース
Infolib-DBR	InfoCom	少	DB統合＋WWW公開 オプション: OAI-PMH , Z39.50
Infolib-ASP	InfoCom	?	同上、ハード不要、アプリケーション部分だけが提供される
iLisSurf e-Lib	富士通	少	電子図書館システム

構築・運用「業者委託 or 自力」

	業者委託	自力
経費負担	大	小
業務負担	小	大
構築スキル	不要	要
カスタマイズ	経済力次第	スキルに応じて自在
その他	業者による担保	システム技術の向上

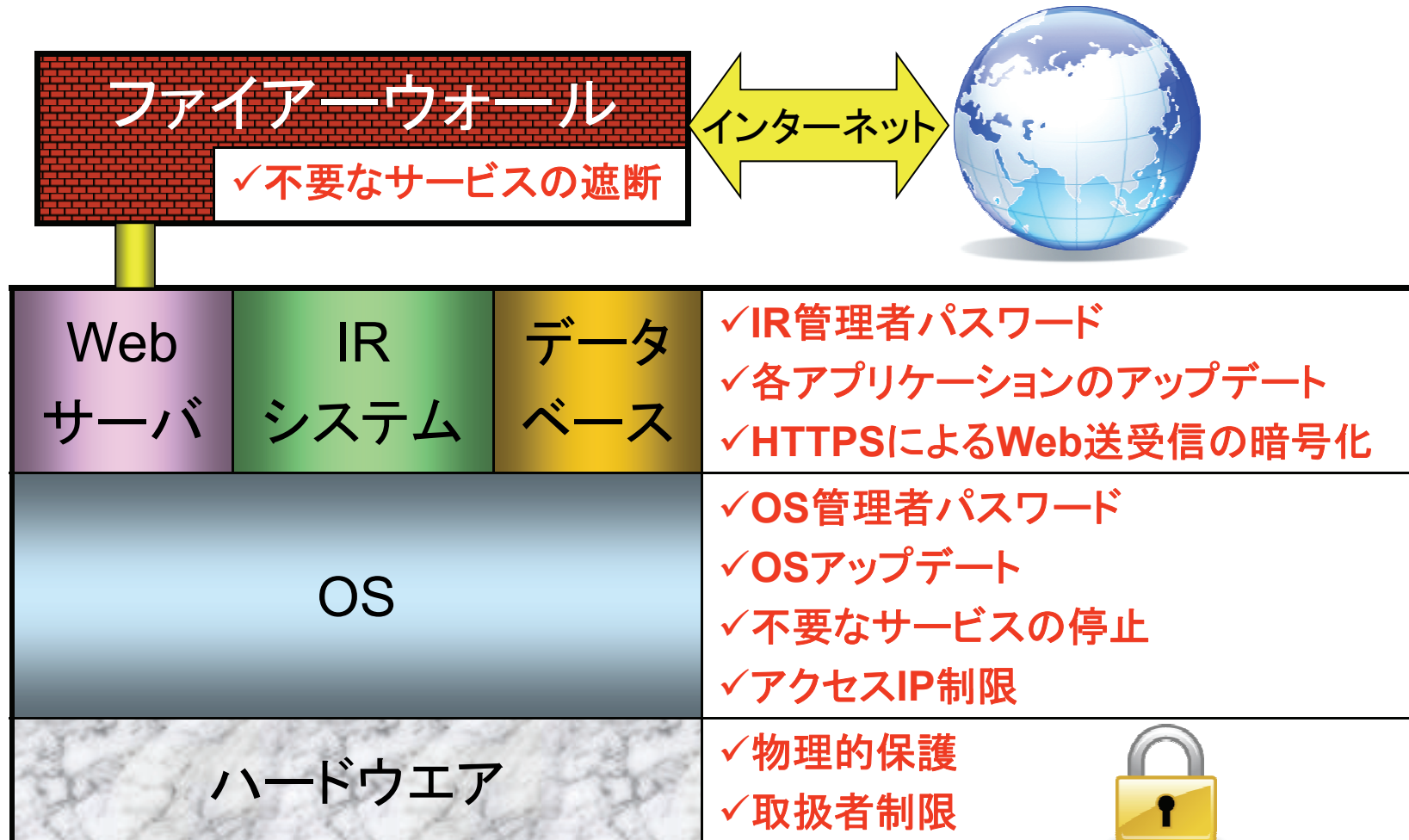
業務項目(部分的業者委託)

- 機器保守
 - 機器保守契約
 - ホスティングサービス
 - 学内システム部門による一括ホスティング
 - 商用ホスティング
 - アプリケーションサービスプロバイダ方式
- カスタマイズなど
 - 機能強化・機能追加
 - ページデザイン

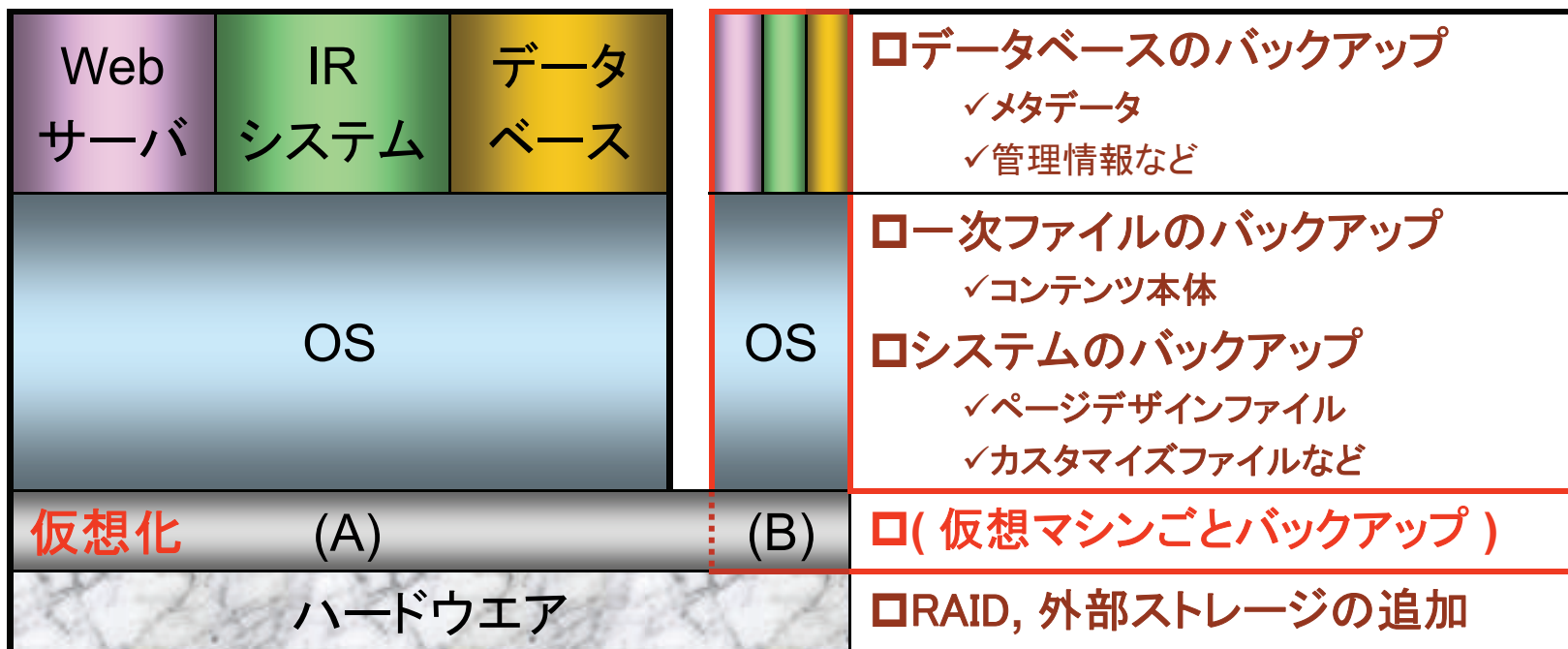
業務項目(部分的業者委託)

- データベース保守
 - ODBCと呼ばれる仕組みを使えばMicrosoft Accessなどを使って、直接データベース内のデータの一括修正・一括削除等が可能
- セキュリティアップデート
- トラブル対応
- コンサルティング
 - 導入・運用・カスタマイズのサポート

セキュリティ



データのバックアップ



➤ RAID1、RAID5、RAID6・・・ディスクの信頼性の向上

–間違って消えてしまったものは戻らない

➤ バックアップ・・・データの複製保存

–ある時点の複製を別の場所に保存・・・テープ、外部ハードディスク、ストレージサービス

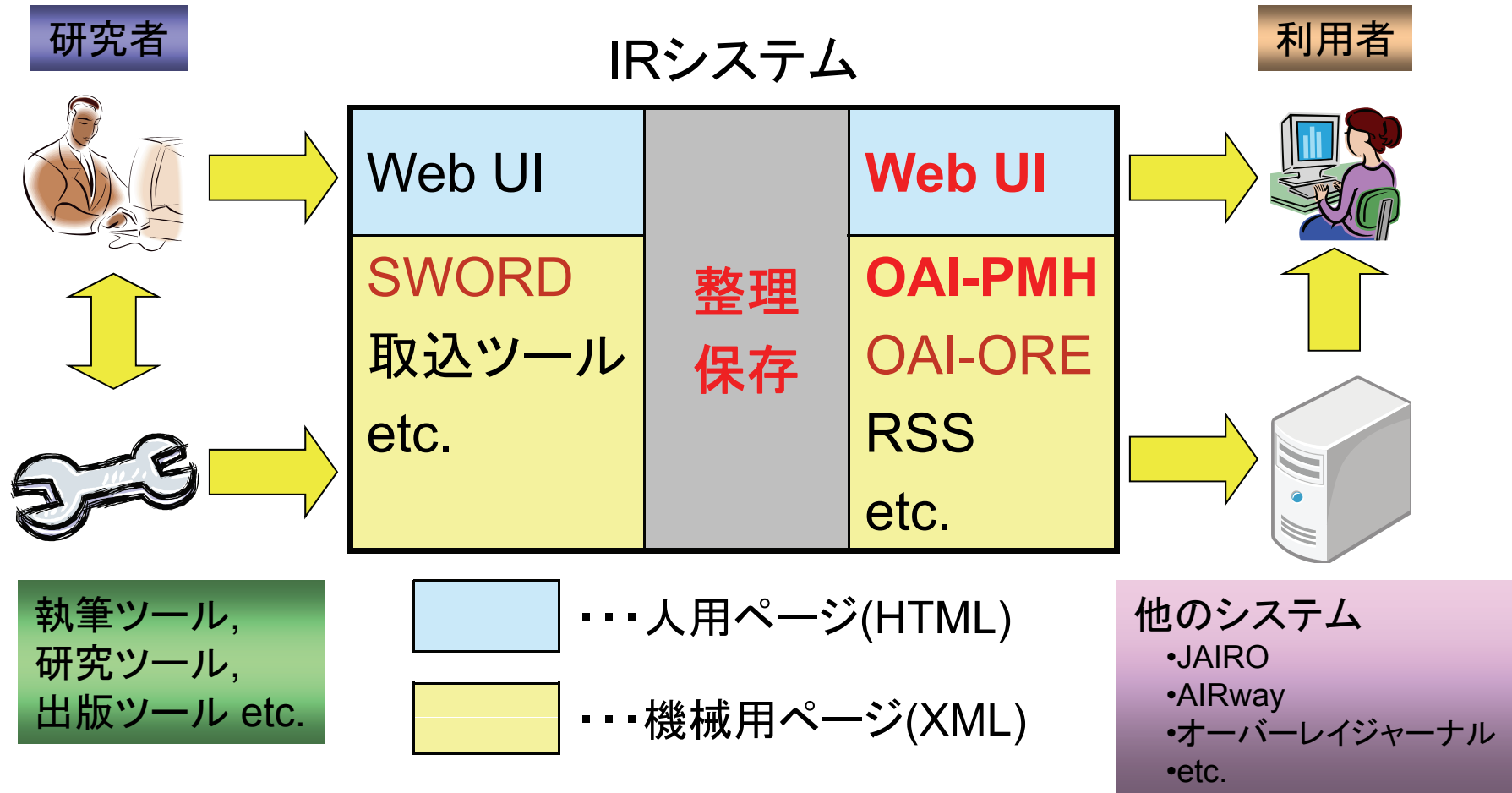
IRシステムの機能(評価ポイント)

- 日本語対応
 - 外観・日本語検索
- 利用統計
 - どのコンテンツがどれだけ閲覧・ダウンロードされたか
- 統合認証
 - IRへの認証を、学内の認証システムやUPKIの認証により行う
 - サービスごとに一々認証の手続をしなくて良いので便利(シングル・サインオン)
- RSS
 - 新着情報の配信
- OAI-PMH
 - システム同士で、メタデータをやり取りする仕組みの一つ
 - IRに必須なのはメタデータを送る側の仕組み(プロバイダ機能)

IRシステムの先進機能(評価ポイント)

- SWORD・・・コンテンツの取り込みの為の標準的な仕組み
- OAI-ORE・・・メタデータ利用のための標準的な仕組み
- コンテンツの出入り口の仕組みはこれまでも色々あったが、
 - (OAI-PMHを除いて)各システムごとに独自のものがほとんど
 - 標準として定められていることは重要
- OAI-PMH並みに、IRシステムの必要条件と呼ばれるものになるかどうかはまだこれからだが・・・
- 時代が進むほどにさまざまな新標準が出てくるのはインターネットの常識
 - シェアや開発組織の状況(活発さ)も評価ポイント

IRシステム機能概要図



DSpace Ver.1.5

- 現行最新バージョン・・・1.5.2
- 日本語対応
 - 外観はデフォルトで多言語対応
 - 日本語のメタデータ検索
 - アスタリスク(*)などを使って日本語の単語をヒットさせることは可能
 - 単語に分割したキーワードでヒットさせるにはカスタマイズが必要
 - 日本語コンテンツの全文検索・・・要カスタマイズ
 - DSpaceで使用されるPDF関連のJavaのライブラリ(機械向けのソフト)が日本語に未対応
 - PerlのPDF関連ライブラリ(xpdf)を使用する設定も可能らしい
 - 少なくともVer.1.4ではコミュニティー名等の多言語化は困難

DSpace Ver.1.5

- 統計機能・・・標準で装備
 - アイテムの閲覧回数など
- 認証
 - 外部認証をカスタマイズできる設計
 - 複数の認証方式の並存が可能
- RSS・・・標準で装備
- OAI-PMH
 - Setの指定はコレクション単位
- OAI-ORE
 - 標準での装備は今のところなし

DSpace Ver.1.5

- SWORD・・・標準で対応
 - SIP(メタデータとコンテンツ本体をまとめたもの)の形式
 - METSDSpaceSIP
- Manakin
 - ページデザインの自由度の向上
- OpenURL
 - 標準装備

DSpace Ver.1.5

- CNRIハンドルシステム
 - 「1234/56」といったハンドルと呼ばれる文字列で、Web上のコンテンツを特定することでそのコンテンツのURLをCNRIが永続的に保証し、サイトの住所変更などによるリンク切れが発生しない
 - でも、ハンドルと、それが示すコンテンツの関連付けを行うのは、コンテンツの管理者側
 - DSpaceを使い続けるなら良いが、他のIRシステムに乗り換える場合のことを考えると、メタデータにハンドルを持たせるとかの細工が必要なはず

DSpace Ver.1.5

- Lightweight Network Interface
 - 情報パッケージ(コンテンツ本体+メタデータ)の取得
 - 情報パッケージの投入
 - etc.

DSpace Ver.1.5

- インストールしてみた感想
 - Mavenという開発ツールが採用され、Ver.1.4のときと違う雰囲気
 - Mavenにより開発がより便利になるらしい
 - DSpaceのインストール自体はVer.1.4のときより簡単
 - Tomcatのセキュリティー関連ではまる
 - 多くのアプリケーションが、セキュリティーが強化された設定を標準にする傾向がある

DSpace Ver.1.4

- 旧バージョン
- 国内ではシェアが飛びぬけて多い
 - 構築・カスタマイズ・運用に関する日本語の情報が多い
 - 業績DB連携、カバーページ、研究者ページ、etc.
 - 日本語化参考サイト
<http://www12.ocn.ne.jp/~zuki/Japanization/index.html>

WEKO Ver.1.3

- 現行最新バージョン・・・1.3.6
 - デモサイト <http://weko.at.nii.ac.jp/demo/>
- 日本語対応・・・日本製なので当然
 - 英語、中国語にも対応
- 統計機能・・・標準装備
 - ダウンロード・ランキングなど
- 認証
 - シングルサインオンの技術の一つであるShibbolethに対応済み
 - 解説ドキュメントはまだの模様

WEKO Ver.1.3

- OAI-PMH
 - 標準でjunii2のフォーマットに対応(さすが日本製)
 - リポジトリ運用者のカスタマイズレベルでサクッとアドオンできるまでは至っていないが、内部設計としては他のフォーマットに対応可能
 - 今のところSet指定は出来ないが、カテゴリ・ツリー単位のSet指定に対応の予定
- OAI-ORE
 - 今のところINDEXツリーについての情報のみ
 - 今後はツリーの終端ノードなどの情報にも対応予定

WEKO Ver.1.3

- SWORD
 - 扱える形式(SIP)はWEKOのエクспорт/インポート形式のみ
 - WEKO用のSWORDクライアント有り
 - ボタン一発で、エクセルファイルからインポート形式を生成してSWORDでWEKOにデポジット
 - 学会等での利用実績有り
- WEKOExport
 - メタデータやコンテンツを他のシステムに出力
 - その他いくつかのExport形式に対応

WEKO Ver.1.3

- NetCommons2のモジュールの1つとして稼動
 - NetCommons2 とは・・・NetCommons2公式ページより
NetCommonsはCMS (Contents Management System)とLMS (Learning Management System)とグループウェアを統合したコミュニティウェアです。NetCommonsによって簡単に構築できるサイトとして、次のようなものが挙げられます。
 - e-ラーニングサイト
 - NPOやNGOのためのバーチャルオフィス
 - 共同研究・学会活動のポータルサイト兼グループウェア
 - オフィス用グループウェア
 - ソーシャルネットワークサービス
 - NetCommons2の機能によりページデザインが簡単
 - 他のモジュールをページに組み込むことが可能
 - IRの新しい使い方につながるかも

WEKO Ver.1.3

- インストールしてみた感想
 - NetCommons2のインストール
 - NetCommons2のインストール解説ページには「Apache(2.0.xx) MySQL4.0.xx PHP4.3.9以降」が必要との説明だけ
 - PHPからMysqlに接続するためのソフトを入れなかったのが原因でハマル
 - WEKOのインストール
 - NetCommons2のインストール解説ページには、「特に以下の組み合わせをお勧めしています(PHP 4.3.9, MySQL 4.1.20, Apache 2.0.52)」とある
 - お勧めバージョンをインストールした結果ハマル
 - WEKOを稼働させるためには、PHPは5系、MySQLは5系が必要
 - ドキュメントや、情報が充実すれば**インストールは簡単**
 - つまずいても、NIIの教員でWEKO開発者である山地先生にメールで質問するととても親切に教えてもらった。

EPrints 3.1

- 現行最新バージョン・・・v3.1.3
 - デモサイト <http://demoprints3.eprints.org/>
- 日本語対応
 - 岡山大で実績あり
 - 全文検索も可能
 - **メタデータの多言語対応も可能**
 - 表示言語に応じた言語でメタデータを表示
 - **DSpaceではできない**
- 統計機能・・・標準では無し、次期バージョンで対応
- RSS, Atom・・・標準で対応
 - 検索結果を、RSS, Atomで配信することが可能

EPrints 3.1

- Plugin
 - 簡単に機能を追加する仕組み
 - メタデータのアウトプット機能, メタデータ取込機能, ユーザーインターフェースのカスタマイズ機能なども Pluginで対応
 - <http://files.eprints.org/view/type/plugin.html>で、各種 Pluginが公開されている
- OAI-PMH
 - junii2の対応はPluginの追加で行う
 - Set指定・・・主題ツリーの各ノードやTypeごとに指定可能
- OAI-ORE・・・v3.2で対応予定？
- SWORD
 - Pluginで対応

EPrints インストール

- OSの種類ごとのインストール方法解説ページ
 - <http://wiki.eprints.org/w/Category:Installation>
- 「[Installing EPrints 3 via apt \(Debian/Ubuntu\)](#)」の場合で試してみた
 - 相当数のPerlのモジュールを個別にインストールしなければならなかった2.3の頃に比べると格段に容易
 - データベースのインストール (コマンド一行)
 - EPrintsのインストール (コマンド一行)
 - サイト名、管理者、DB名、DBパスワード他いくつかの項目を指示に従って設定
 - Apacheの設定と再起動

E-Repository

- 日本語対応
 - メタデータの日本語検索可能(多言語も可)
 - 漢字統合インデックスも
 - 日本語全文検索機能はなし
- 統計機能・・・標準機能(AWstats)
 - <http://mitizane.ll.chiba-u.jp/metadb/statics/awstats.html>
- RSS・・・標準機能
 - 更新日を元に行っているため、修正でも出てしまうのが難

E-Repository

- OAI-PMH
 - Set指定は2つまで
 - <http://mitizane.ll.chiba-u.jp/curator/oai.html>
- インポート・ツール
 - テキストファイル(csv)による一括登録
- エクスポート・ツール・・・標準機能(csv)
- OpenURL対応済み

XooNips Ver.3.4

- 最新バージョン3.4.2a
- 日本語対応
 - メタデータ、本文の全文検索に対応
- 統計機能
 - 月別のダウンロード数、アクセス数、ランキングなど、
 - 個々のメタデータにもダウンロード数や閲覧数が表示される
- OAI-PMH
 - インデックスツリーごとにSetの指定が可能

XooNips Ver.3.4

- インポート・ツール
 - EXCEL,CSVファイル+本文ファイルから、インポートファイルを生成するツールが提供され、一括インポートが可能
- エクスポート・ツール
 - XML形式でメタデータを選択的または一括エクスポートする機能有り
- ある使用者の実感
 - 事実上、唯一の国産オープンソースによるリポジトリソフトウェアで開発元が日本国内なので、開発元と直接コミュニケーションを取ったり、要望が出すことが容易
 - XOOOPSをベースにしているので、拡張が容易
 - DSpaceが大勢を占めている日本の現状では、機関リポジトリのための新しい開発がDSpaceベースで進められてしまうため、流れに乗り遅れてしまう傾向がある

最後に

- システムよりコンテンツが大事
 - システム構築にかけるお金はなるべく節約して
 - コンテンツ収集に注力する
- 機関リポジトリとは何か？
 - IRシステムには新しい機能がどんどん開発・追加されている
 - IRに関する思想はシステムの機能として反映されている
 - リポジトリをシステムレベルで理解することの良さ大切さ
 - システム構築について、ハードルは下がり、情報は増える傾向にあるので、試しにインストールしてみてもは？