平成20年度学術ポータル担当者研修 機関リポジトリシステム概論

NII学術ネットワーク研究開発センター 山地 一禎 2008-08-28

内容

- ▶ システム概略
 - リポジトリシステムとは?
 - システム構成
- リポジトリシステム
 - DSpace
 - EPrints
 - WEKO
- メタデータ流通
 - OAI-PMH
 - OAI-ORE

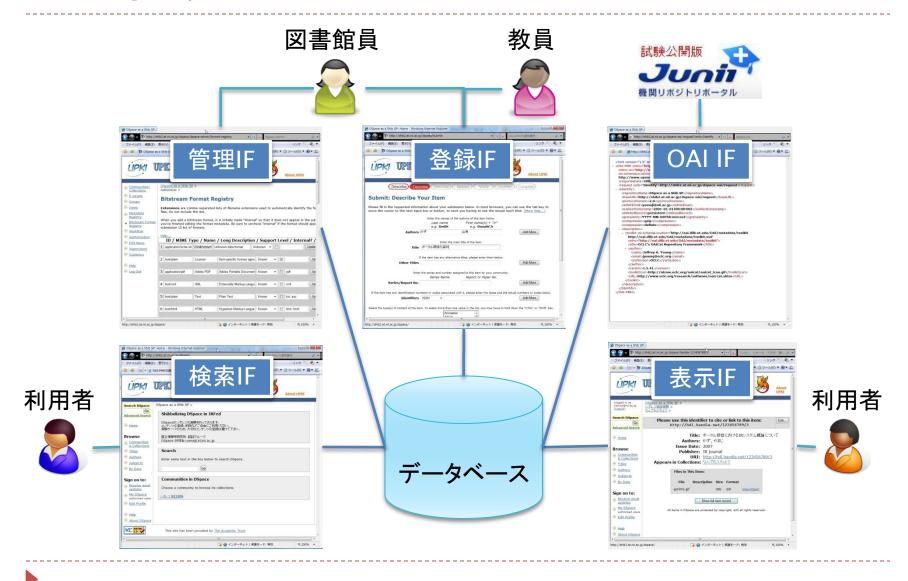
目標

- リポジトリシステムがやってることを理解してもらうこと
- 自分でリポジトリを立てようと思う人を増やすこと

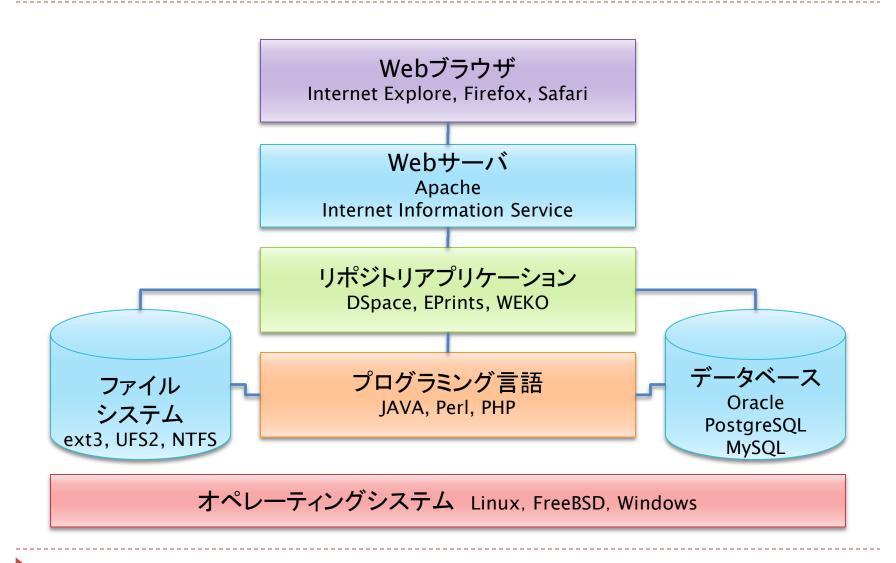
システム概略 必要な機能

- ▶ Webインターフェースを提供すること
- メタデータとファイルが管理できること
- ▶ OAI-PMHに対応していること

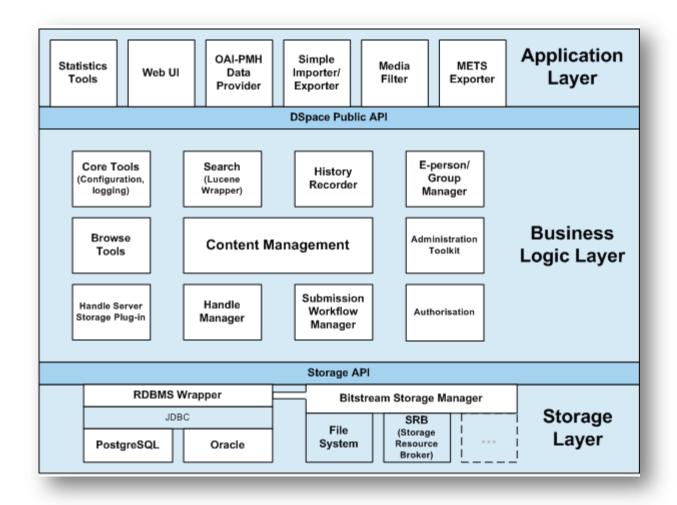
システム概略 Webインターフェース



システム概略システム構成



システム概略 DSpace Architecture



リポジトリシステム

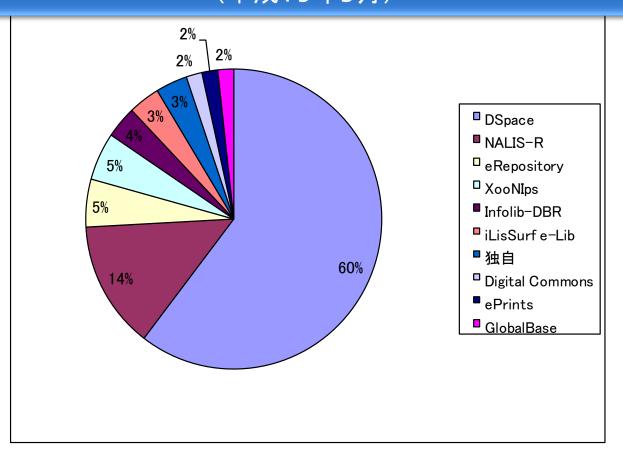
一般的な選択肢

製品種類	構築方式	必要スキル	金銭コスト
オープンソース	自力	高	低
	導入支援	低~中	中
	パッケージ	低~中	高
商業製品	ローカルホスト	低~中	高
	ASP	低~中	中
独自システム	自力	高	低~高
	開発·導入委託	低~中	高

- サポート(カスタマイズ追加,不具合対策,バージョンアップ等)も要考慮
- 図書館だけではなく、学内情報関連組織との連携が望まれる

リポジトリシステム 国内シェア

「次世代学術コンテンツ基盤共同構築事業(中間まとめ)」 (平成19年3月)



リポジトリシステム 商用製品

- ローカルホスト式
 - ▶ Infolib-DBR (国産:InfoCom)
 - http://www.infocom.co.jp/das/infolib/dbr/index.html
 - ▶ E-repository (国産:CMS/NEC)
 - http://www.cmsc.co.jp/modules/office/index.php?id=12
- ▶ Application Service Provider式
 - ▶ Infolib-ASP (Infolib-DBRのASP版)
 - http://www.infocom.co.jp/das/asp/index.html
 - ▶ Digital Commons (米:Berkeley Electronic Press)
 - http://www.bepress.com/ir/

リポジトリシステム オープンソース

DSpace

- ▶ 米国·MIT, HP開発
- JAVA, PostgreSQL or Oracle
- デモ http://shib2.at.nii.ac.jp/dspace/

EPrints

- 英国・サウサンプトン大学開発
- Perl, MySQL
- デモ http://eprints.at.nii.ac.jp/

WEKO

- ▶ 日本・国立情報学研究所開発
- PHP, MySQL
- デモ http://weko.at.nii.ac.jp/









リポジトリシステム DSpace

- DSpace系で良くやっている機能強化
 - メッセージの日本語化
 - ▶ メタデータ検索の日本語対応化
 - 全文検索の日本語対応化
 - ▶ junii2対応
 - 詳細な利用統計
 - ハンドル

やろうと思えば 無料でできる

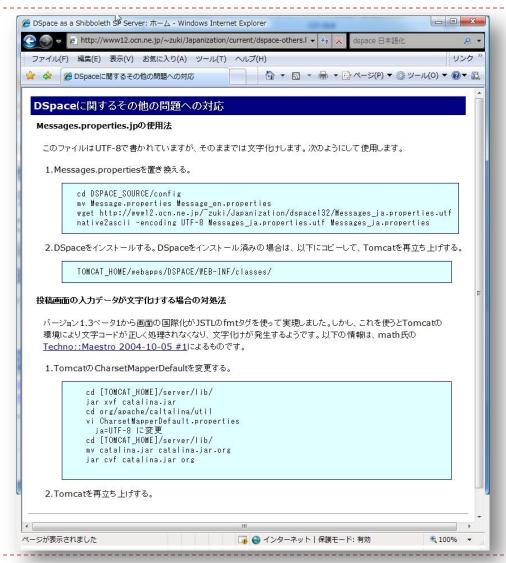
DSpaceの日本語化について

http://www12.ocn.ne.jp/~zuki/Japanization/index.html http://repository.kulib.kyoto-u.ac.jp/dspace-memo/japanization.html

リポジトリシステム DSpaceメニューの日本語化



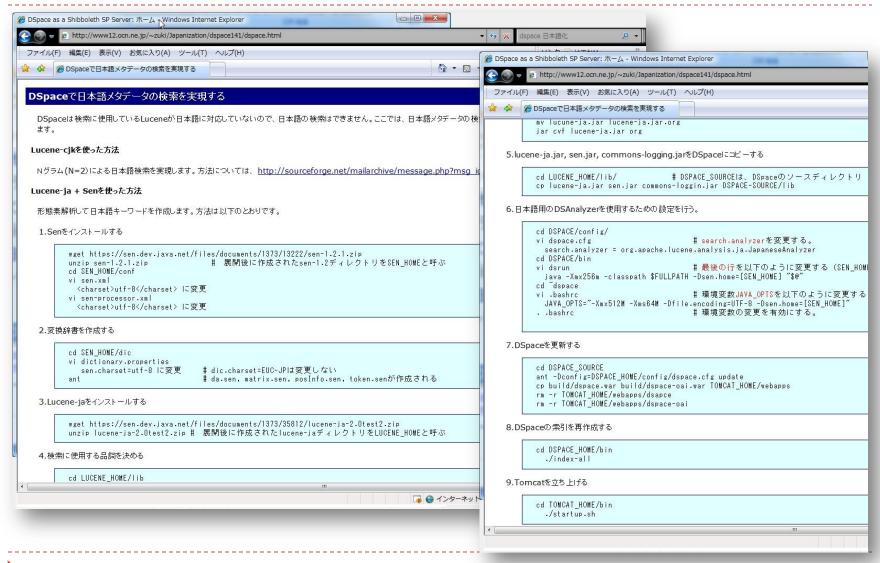
リポジトリシステム DSpaceメニューの日本語化



リポジトリシステム DSpaceメタデータ検索の日本語化



リポジトリシステム DSpaceメタデータ検索の日本語化



リポジトリシステム DSpace お茶の水女子大学TeaPotサイト



リポジトリシステム DSpace 1.5

DSpace 1.5 Overview





- DSpace "modules"
- Application "Overlays"
- XMLUI (aka. Manakin)



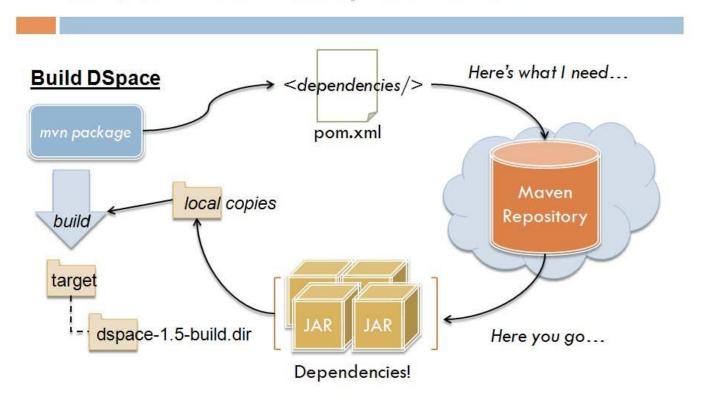
- Configurable Submission
- □ Streamlined Browse System
- □ SWORD Interface



LNI (Lightweight Network Interface)

リポジトリシステム DSpace 1.5 Maven

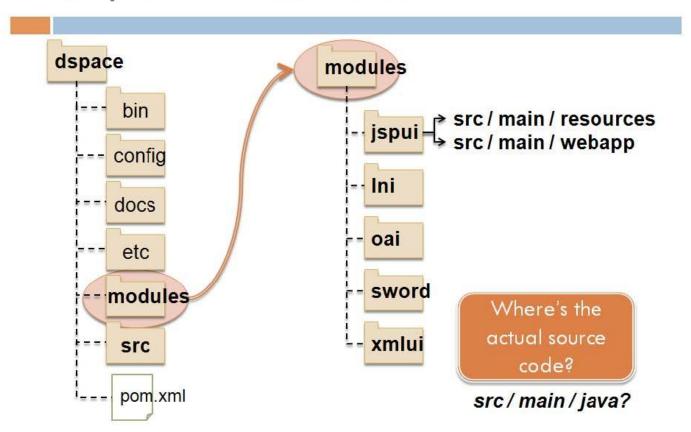
Ok, so what really is Maven?



Donohue, T. and Triggs, G. (2008) *Customizing DSpace 1.5 with Basic Overlays.* In: Third International Conference on Open Repositories 2008, 1–4 April 2008, Southampton, United Kingdom.

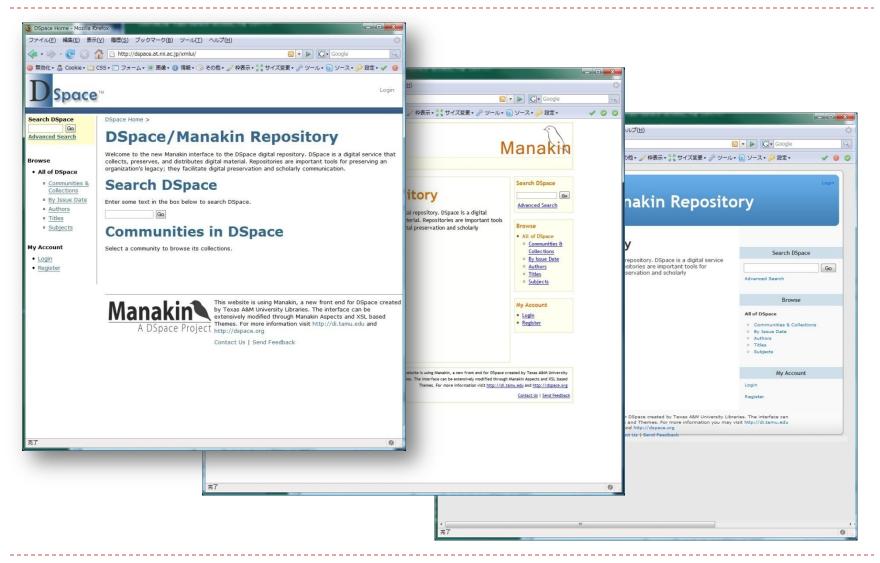
リポジトリシステム DSpace 1.5 Maven

DSpace 1.5 "Modules"

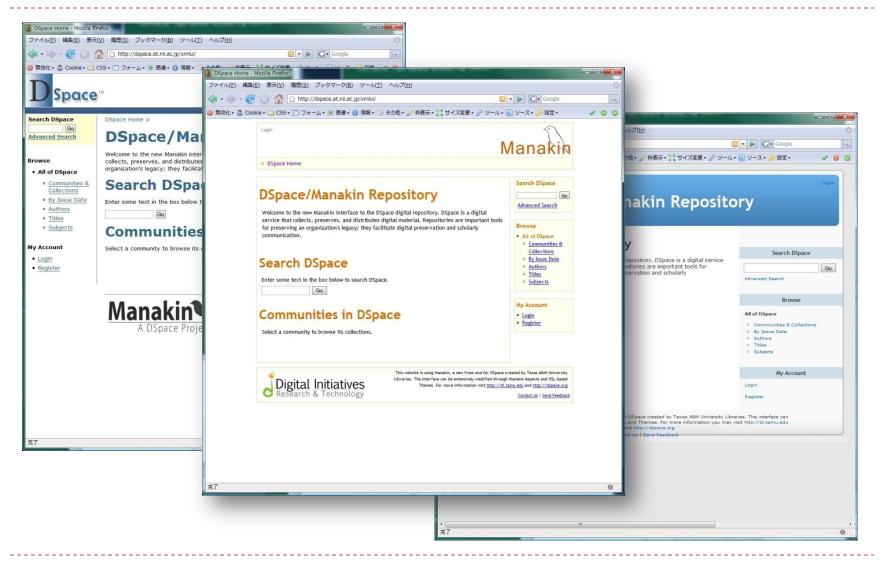


Donohue, T. and Triggs, G. (2008) *Customizing DSpace 1.5 with Basic Overlays.* In: Third International Conference on Open Repositories 2008, 1-4 April 2008, Southampton, United Kingdom.

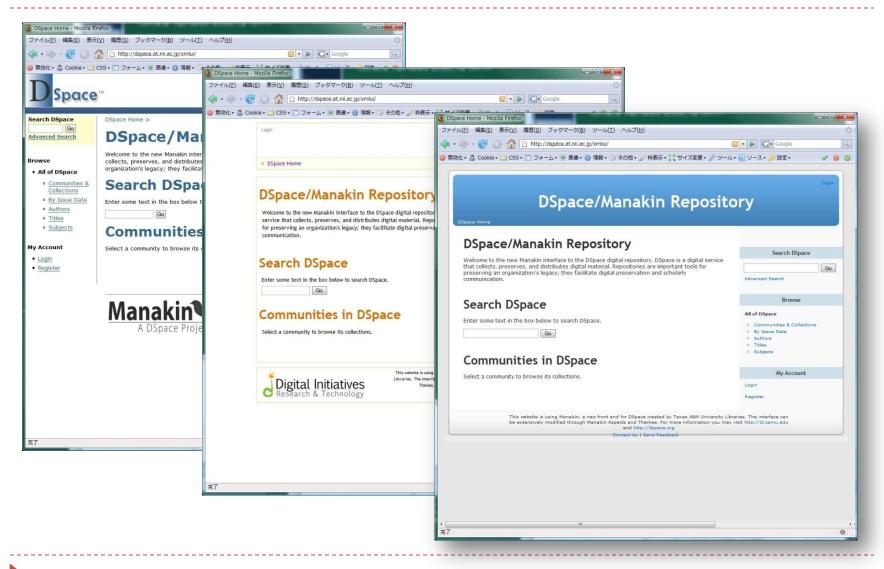
リポジトリシステム DSpace 1.5 Manakin



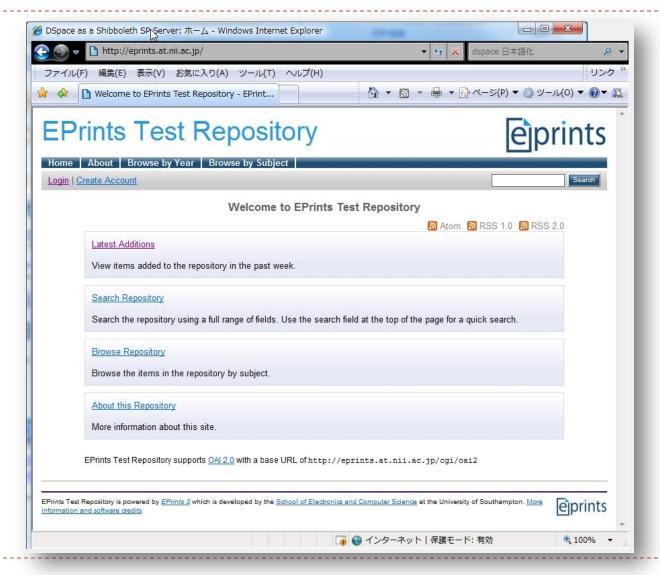
リポジトリシステム DSpace 1.5 Manakin



リポジトリシステム DSpace 1.5 Manakin



リポジトリシステム EPrints デモサイト



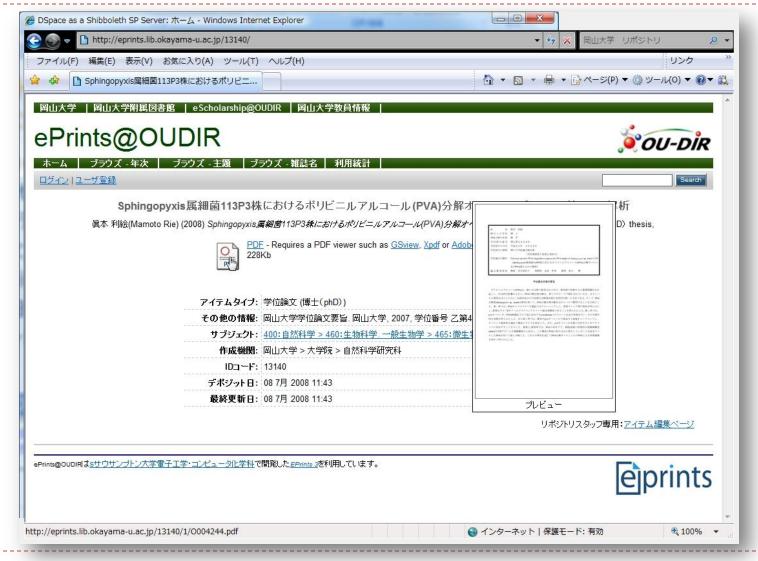
リポジトリシステム EPrints OR08サイト



リポジトリシステム EPrints 岡山大学OU-DIRサイト



リポジトリシステム EPrints 岡山大学OU-DIRサイト



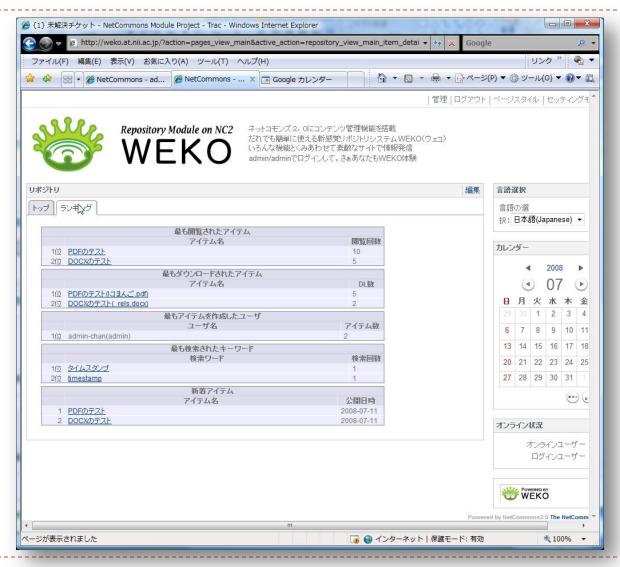
リポジトリシステム WEKO



リポジトリシステム WEKO一般編 アイテム詳細



リポジトリシステム WEKO一般編 ランキング



リポジトリシステム WEKOログインユーザ編 アイテム登録



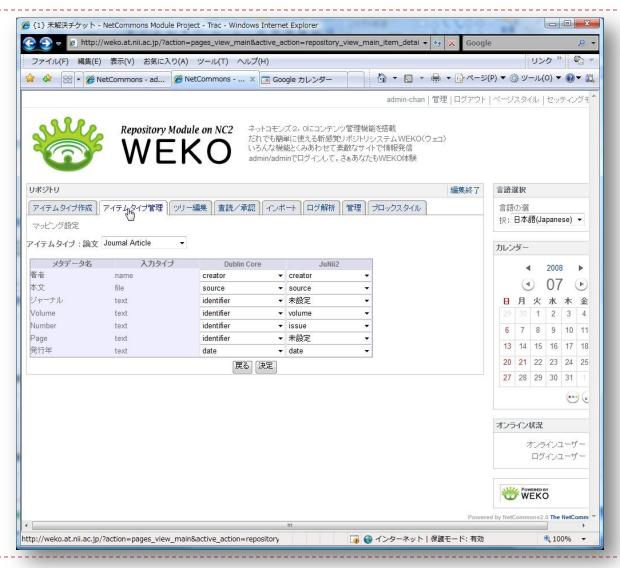
リポジトリシステム WEKOログインユーザ編 ワークフロー



リポジトリシステム WEKO管理編 アイテムタイプ作成



リポジトリシステム WEKO管理編 アイテムタイプ管理



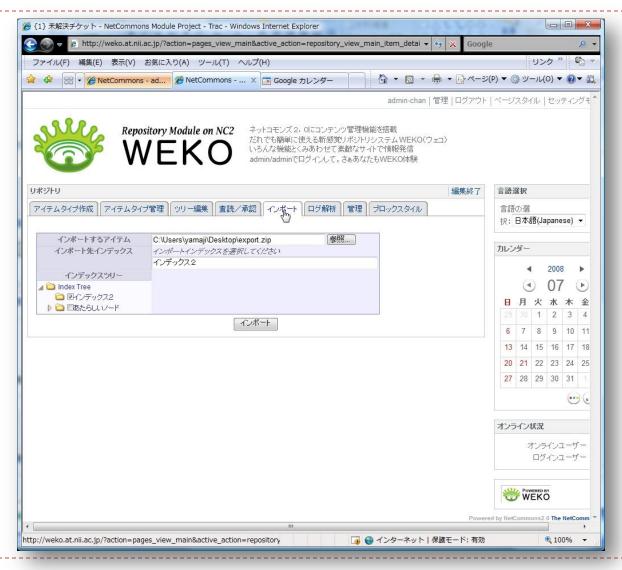
リポジトリシステム WEKO管理編 ツリー編集



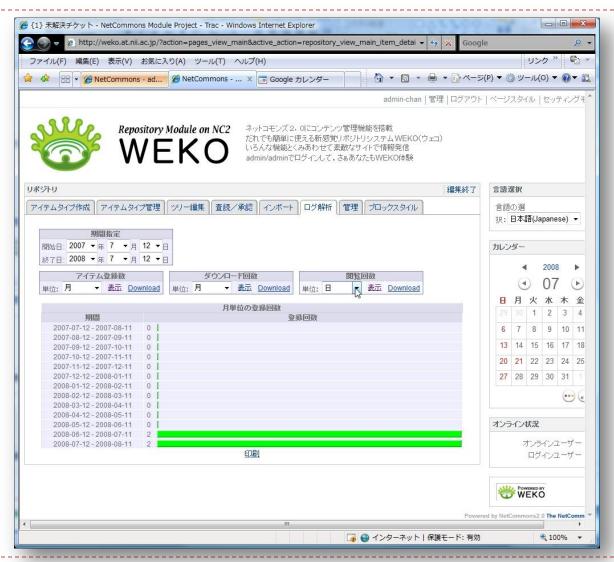
リポジトリシステム WEKO管理編 査読・承認



リポジトリシステム WEKO管理編 インポート



リポジトリシステム WEKO管理編 ログ解析



リポジトリシステム WEKO管理編 管理



リポジトリシステム WEKO カレント



OAI-PMH リポジトリ構築後のステップ



OAI-PMH リポジトリ構築後のステップ:ついでに



OAI-PMH 登録データの流通

登録したデータをどのようにして広めていくか?

- · Googleの検索対象になるように努力
- データベース間で明示的にネットワークを確立

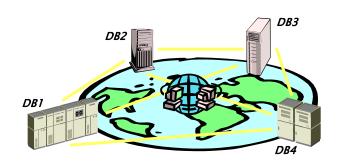
OAI-PMH

(Open Archive Initiative - Protocol for Metadata Harvesting)

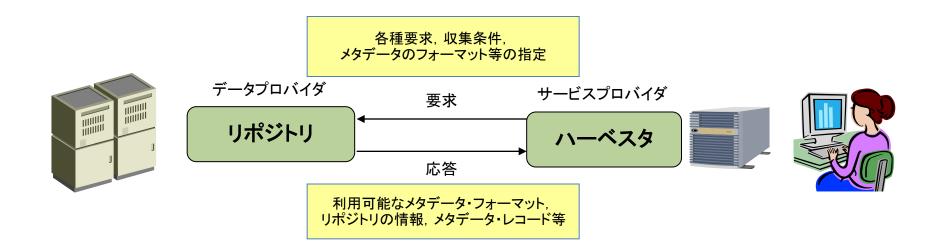
サーバ間でメタデータを機械的に収集するための、要求と結果を授受するためのプロトコル.

1999年10月: サンタフェ会議でFrameworkが合意

2001年 1月:OAI-PMH 第1版 2002年 6月:OAI-PMH 第2版



OAI-PMH メタデータ通信



Dublin Coreメタデータセットでの交換を必須

Title	Creator	Subject	Description	Publisher
Contributor	Date	Type	Format	Identifier
Source	Language	Relation	Coverage	Rights

OAI-PMH junii2フォーマット①

・タデータフォーマット(juni2) 各データ要素の入力内容一覧

凡例: 青二多項、黄色=あれば多項 左記以外=各機関リポジトリのポリシーによる

32	大項目	項目	element	内容	6 3	構造し可否	minOccurs	max0cours	書式(空白の場合は文字列)	oal_doにダムダウンさせる ときの行き先	オブションの責性
1		タイトル	title		63	×	1	1		title	lang
2	タイトル	その他(別書語等)のタイトル	alternative	別書語のタイトルやタイトルのヨミ			0	unbounded		title	lang
3		作成者	creator		あれば必須		0	unbounded		creator	lang
4		着者キーワード	subject				0	unbounded		subject	
5		国立情報学研究所メタデータ主題語彙集	Nileubject	下記URL参照 http://www.nlac.jp/metadata/manual/NE-category.pdf			0	unbounded		subject	version
ê		日本十進分難法	NDC				0	unbounded		subject	version
7		国立国会园書館分類表	NDLC				0	unbounded		subject	version
8	土班	日本件名標目	BSH				0	unbounded		subject	version
9	1.2	国立国会国書館件名揮目表	NDLSH				0	unbounded		subject	version
10		医学件名標目表	MeSH				0	unbounded		subject	version
11		デューイナ連分類法	DDC				0	unbounded		subject	version
12		米国議会図書館分類表	LCC				0	unbounded		subject	version
13		国際十進分競法	UDC				0	unbounded		subject	version
14		米国議会図書館件名揮目表	LCSH				0	unbounded		subject	version
15		内容配送	description	論文の抄録 学校論文にあっては、授与年度、論文/課程の種別等 会議発表論文にあっては、会議名、開始地、日時等 その他コンテンツの内容を示すことを記述			0	unbounded		description	
16		公開者	publisher	コンテンツ本体の公開を行った個人、組織、団体			0	unbounded		publisher	lang
17		寄与者	contributor	コンテンツ本体の作成に関わりを持つ、個人、組織、団体			0	unbounded		contributor	lang
18		日付	date	コンテンツの作成日付			0	unbounded	YYYY-MM-DD, YYYY-MM, YYYY	date	
19		資源タイプ	type	コンテンツの種類(自由記述)			0	unbounded		type	
20		国立情報学研究所 メクデーラ主雑語条集 (資語タイプ)	Nilhype		6 ∰	×	1	1	[Journal Article Thesis or Dissertation Departmental Bulletin Paper Conference Paper Presentation Book Technical Report Research Paper Article Preprint Learning Material Date or Dataset Software Others	туре	
21		フォーマット	format	Internet Media Typeで規定されるファイル形式			0	unbounded		format	
22		その他の資源識別子	identifier	23項、24項以外の資源識別子			0	unbounded		identifier	
	資源 職別子	資源練別子URI (アイテム表示画面へのリンク)	URL	コンテンツ本体を表示する関面へのURL	6 %	×	1	1	URL	identifier	
24		本文フルテキストへのリンク	fullTextURL	コンテンツ本体へのURL	あれば必須		0	unbounded	URE	identifier	
25		ISSN	issn		あれば必須		0	unbounded	Wd[4]W-7Wd[3][WdXx]		
26		書誌レコードID(総合目録DB)	NCID		あれば必須		0	unbounded		折りたたみ編集の上、	
27		雑誌名	ititle	本 タイトルのみ	あれば必須		0	1			lang
28	再數數	•	volume		あれば必須		0	1			
29	情報	*	issue		あれば必須		0	1		identifier	
		開始ページ	spage	論文の開始ページ	あれば必須		0	1	半角数字		
30				株本の株では、20	あれば必須		n .	1	半角数字		
30		終了ページ	epage	論文の終了ページ	404 F14-80-38						
30 31 32		教了ページ 刊行年月	dateofissued	無文の表 [ペーン	あれば必須		0	1	YYYY-MM-DD, YYYY-MM, YYYY		

OAI-PMH junii2フォーマット②

温書	大項目	現 日	element	内容	多項	機道し可否	minOcours	max0ccurs	書式(空白の場合は文字列)	cal_doにダムダウンさせる ときの行き先	オブションの異性
34	-	##	language	コンテンツ本文の言語			0	unbounded	ia, jon. ja JP	language	
35		P4 80	relation			-	0	unbounded		relation	
38		D.Mad St.	pmid				0			relation	
37	ID		doi			-	0			relation	
38			ls/VersionOf	当該コンテンツに対して、参照先のコンテンツを生たる版と みなすことができる。			0		URI	relation	
39	1	異版あり	hasVersion	当該コンテンツが参照先のリソースを別なる版として持つ			0	unbounded	URL	relation	
40	1	世換される	isReplacedBy	当該コンテンツが参照先のリソースによって置き換わる			0	unbounded	URI	relation	
41	1	登換する	replaces	当該コンテンツが参照先のコンテンツを新しく置き換える			0	unbounded	URL	relation	
42		要件とされる	isRequiredBy	●開先のコンテンツを利用するにあたり、当該コンテンツを 必要とする			0	unbounded	URL	relation	
43	版に製 する情	要件とする	requires	当該コンテンツを利用するにあたり、参照先のコンテンツを 必要とする			0	unbounded	URL	relation	
44	•		isPartOf	当該コンテンツが参照先コンテンツの部分をなす			0		URL	relation	
45	7		heePart	当該コンテンツが参照先のコンテンツを部分として持つ			0			relation	
46			isReferencedBy	当該コンテンツが参照際のコンテンツから関連づけられる			0			relation	
47		参照する	references	当該コンテンツが参照先のコンテンツを関連付けている			0	unbounded	URL	relation	
48		別フォーマットである	leFormatOf	当該コンテンツが参照先のコンテンツと内容的に同じである が、Formatは異なる			0	unbounded	URL	relation	
49		別フォーマットあり	hasFormat	当該コンテンツが参照先コンテンツに異なるフォーマットを 持つ			0	unbounded	URL	relation	
50			coverage				0	unbounded		coverage	
51	地理的	空間的	spetial				0	unbounded		coverage	
52	新 田		Nilspatial				0	unbounded		coverage	
53			temporal				0	unbounded		coverage	
54	時系列 範囲		Niltemporal				0	unbounded		coverage	
55		権利	rights	コンテンツ本体に関する権利規定			0	unbounded		rights	
58		著者版フラゲ	textversion	コンテンツ本体のパージョン author 著者様 publisher 出版柱版 none:上記以外			0	1	[author publisher none]		

オプション賞性 書式 Bang RFC1766 version 本数

OAI-PMH JuNii2資源タイプ

項目	内容				
Journal Article	学術雑誌論文	学術雑誌に掲載された論文			
Thesis or Dissertation	学位論文	博士論文、修士論文等 授与年度、論文/課程、学位 の種別等は必要に応じてDescription に記入する。			
Departmental Bulletin Paper	紀要論文	紀要類に掲載された論文			
Conference Paper	会議発表論文	会議の報告書等に掲載された論文 会議名、開催地、 日時等は必要に応じてDescriptionに記入する。			
Presentation	会議発表用資料	会議で発表されたプレゼンテーション資料、ポスター、 口頭発表資料等 会議名、開催地、日時等は必要に応 じてDescription に記入する。			
Book	図書	図書の章(Chapter)も含む。			
Technical Report	テクニカルレポート	テクニカルレポート、ディスカッションペーパー、 ワーキングペーパー等の機関発行の報告書			
Research Paper	研究報告書	科研費やCOE等、研究助成金による研究成果の報告書 種別(助成元、研究課題の種類)等は必要に応じて Description に記入する。			
Article	一般雑誌記事	学術論文以外の記事 (コラム等)			
Preprint	プレプリント	プレプリント			
Learning Material	教材	授業等で用いる資料類			
Data or Dataset	データ・データベース	実験記録等のファクトデータ、及びそれらの集合から なるデータベース等			
Software	ソフトウェア	ソフトウェア			
Others	その他	上記以外のもの			

OAI-PMH 要求(verb)一覧

	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		

OAI-PMH 要求例

ベースURL

http://teapot.lib.ocha.ac.jp/ocha-oai/request?

verb=ListRecords

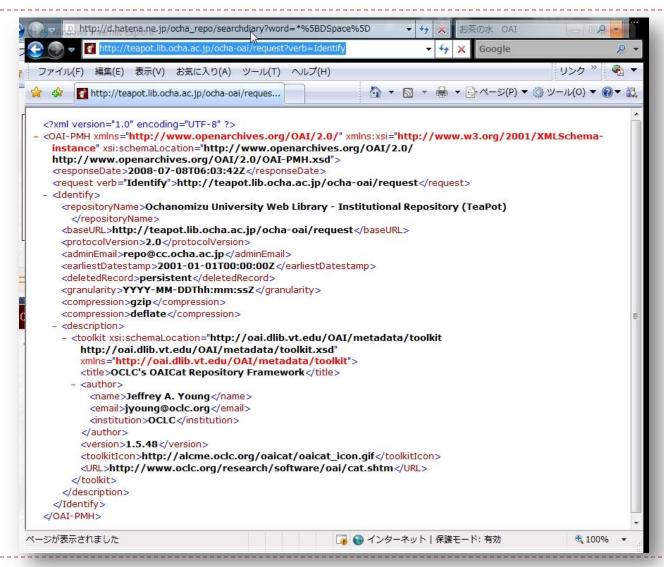
&metadataPrefix=junii2

要求の種別

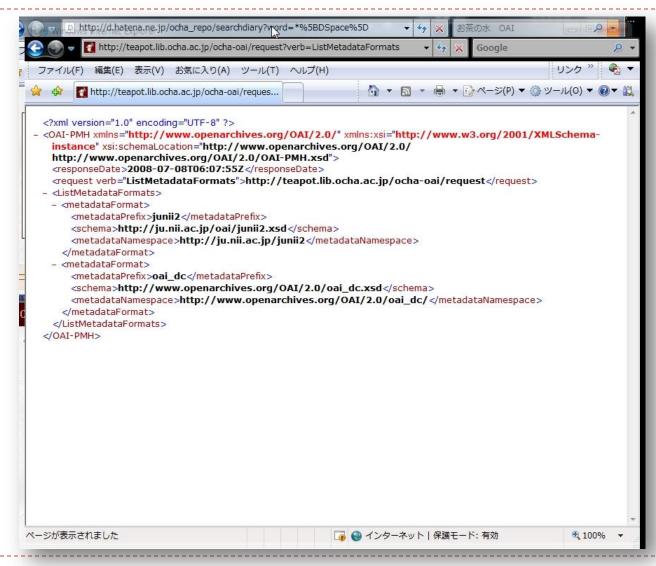
引数

リポジトリ(http://teapot.lib.ocha.ac.jp/ocha-oai/request)から
Junii2メタデータフォーマットで公開されているメタデータを
全て取得せよ

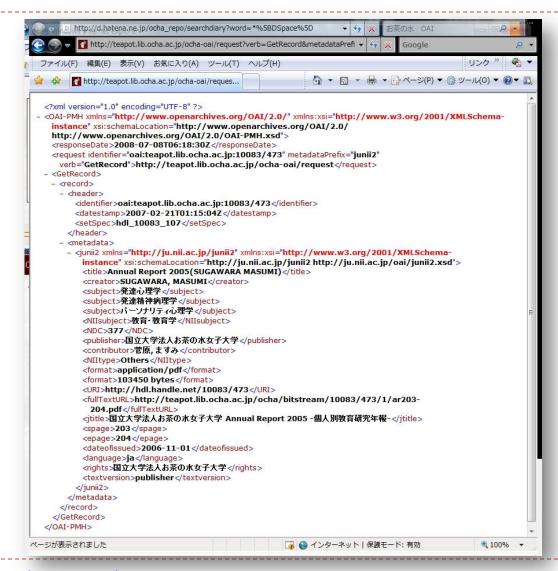
OAI-PMH Identify



OAI-PMH ListMetadataFormats



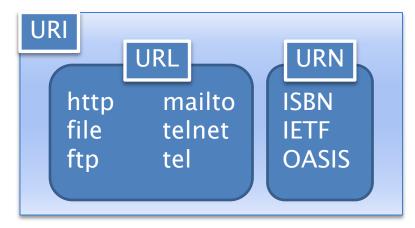
OAI-PMH GetRecord



OAI-PMH

参考情報:情報の固定化

- URI Uniform Resource Identifier
 - ▶ リソースを特定する文字列
- URL Uniform Resource Locator
 - リソースをアクセスする方法で特定
 - http://domain/path , mailto:mbox@domain
 - http://hdl.handle.net/1721.1/29466
- URN Uniform Resource Name
 - ▶ リソースを恒久的な名前で特定
 - urn:isbn:4-8399-1796-5
 - urn:handle:1721.1/29466



- ・大切なのはWeb資源の固定化
- HandleやDOIは単なるURN

OAI-ORE



OAI-ORE 「ORE仕様書とユーザガイド アルファ0.2版 日本語訳」

http://www.nii.ac.jp/irp/archive/translation/oai-ore/0.2/



OAI-ORE どんなものなのか?

インターネット上のリソースの関連付けを明確にして発信

対象

- ▶ 関連付けられたHTMLページのセット
- ソーシャルネットワークサイト内でのリソースセット
 - 本文, 画像, コメント
- リポジトリ内での複数フォーマットからなるリソースセット
 - ▶ LaTeX, PDF, HTML
- オーバーレイジャーナル

▶ 御利益

- サーチエンジンがより賢くクロールできる
- ブラウザがより賢くナビゲートできる
- コンテンツアーカイブのためのシステムがより賢く働く
- 引用関係などの情報をより賢く機械処理できる

学術情報のリュースを可能にし次世代学術コミュニケーションを形成

税金や授業料を何に使うか?

- リポジトリシステムは発展途上
 - ▶ 現在でも新しい機能がどんどん開発・追加されている
 - リポジトリをシステムレベルで理解することの重要性
 - リポジトリとは何かを考えることが重要
 - ▶ リポジトリに関する思想はシステムの機能として反映されている
- システム構築はそれほど難しくない
 - システム導入のハードルは低くなってきた
 - 経験者. 情報サイト. 各種サポートベンダーあり
 - ▶ 試験運用や実験では、パソコンレベルでも可
 - ▶ 本格運用でも多分問題ない
 - ▶ バックアップ装置は要必要
 - ▶ 1万円のUSB-HDDでも十分かも知れない
- 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極力少なく
 - サクッと立ち上げてコンテンツ収集に注力する
 - ▶ リポジトリとは何かをよ~く考える