

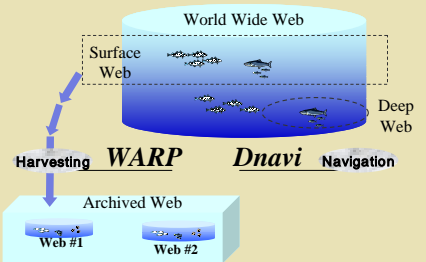
二つのプロジェクト

- ◆ インターネット資源選択的蓄積実験事業 (WARP)
 - 表層ウェブを選択的に収集
 - 納本制度審議会の議論に資する
- ◆ データベース・ナビゲーション・サービス (Dnavi)
 - (主に) 深層ウェブの入口までナビゲートするリンク集

2002/12/14

9

二つのウェブと二つのプロジェクト



2002/12/14

10

経緯と計画

- ◆ 1998 電子図書館構想
- ◆ 1999 電子図書館サービス実施基本計画
- ◆ 2000-2002 プロトタイプシステム開発
- ◆ 2002.1 国際シンポジウム開催
- ◆ 2002.3 納本制度審議会諮問
- ◆ 2002.6-2004 三カ年実験事業
- ◆ 2002.11 WARP、Dnavi公開
- ◆ 2004 納本制度審議会答申(予定)

2002/12/14

11

Dnavi (Database Navigation Service)とは？

- ◆ リンクによるデータベースへの案内
 - データベース = 収集できない深層ウェブの中でも特に有用な情報資源
- ◆ 書誌情報: タイトル、作成者、分類、内容
- ◆ データベースのポータル・サイトとして約5,000件を収録
- ◆ レファレンスツールとしても有効
- ◆ 横断検索ではない

2002/12/14

12





WARP(Web Archiving Project)とは？

- ◆ 著作権者との許諾契約に基づく
- ◆ インターネット上の情報資源を選択的にロボット収集
- ◆ 定期的な再収集・更新管理、時系列で蓄積
- ◆ 更新や削除等によってインターネット上から消滅した後においても、過去の情報へのアクセスを可能とするためのサービス
過去のウェブへ「ワープ」(WARP)

2002/12/14

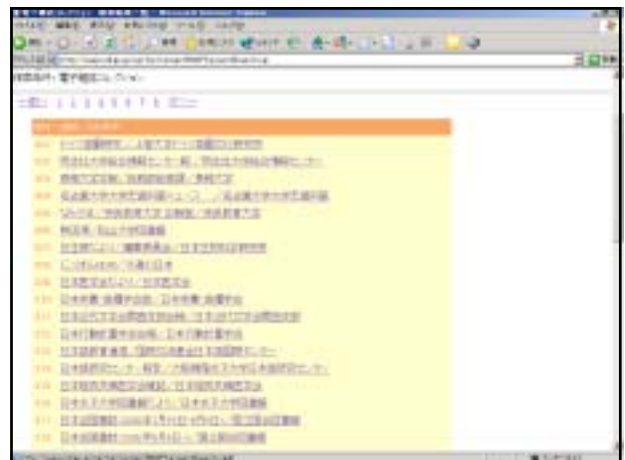
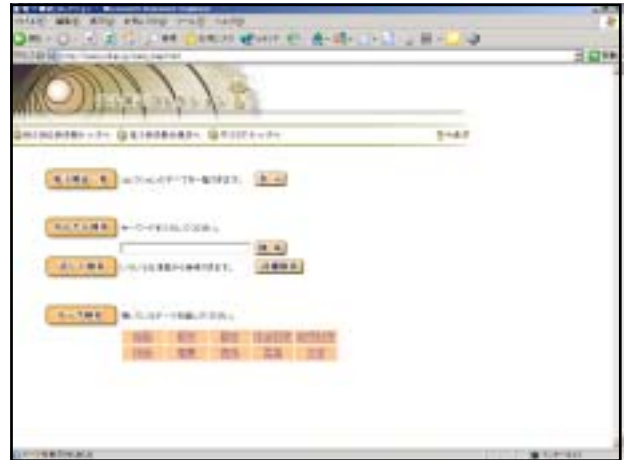
31

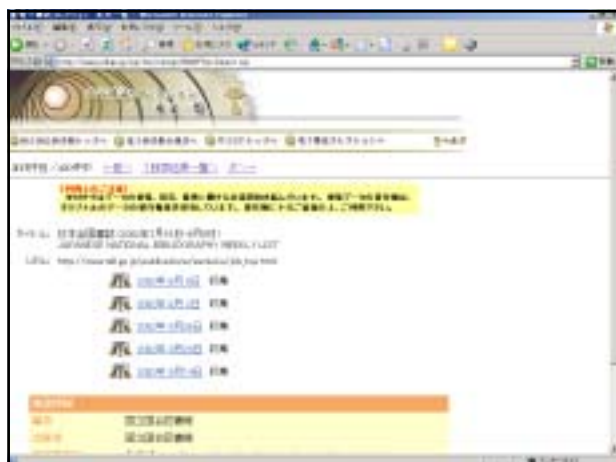
WARPのコレクション

- ◆ 電子雑誌コレクション
– 約400タイトル
- ◆ 政府ウェブコレクション
– 参議院等
- ◆ 協力機関コレクション
– FIFAワールドカップ日本組織委員会
– 2002年「日本年」「中国年」実行委員会

2002/12/14

32



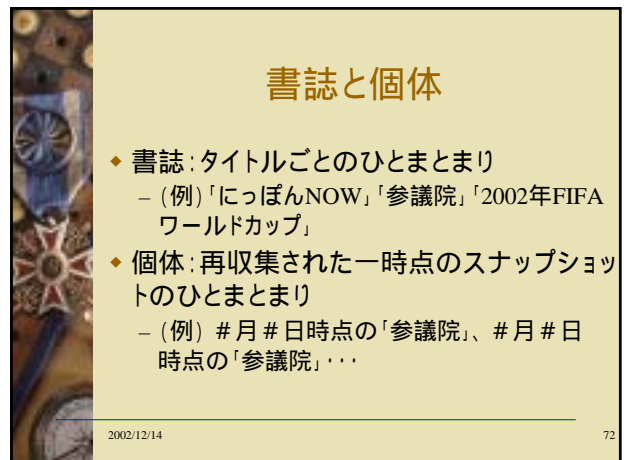


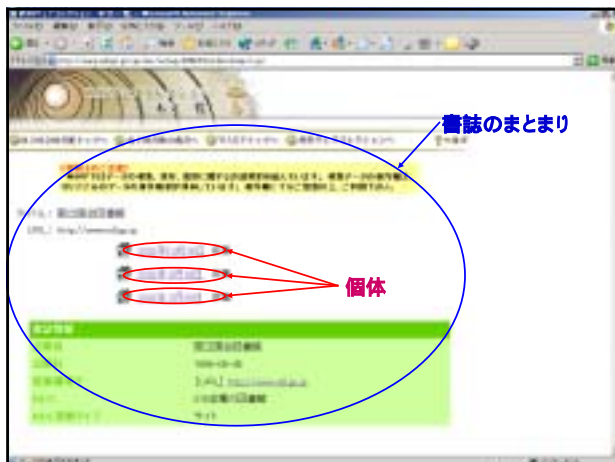












WARPの業務モデル

- ◆ 新規収集
 - 収集対象の発見
 - 調査
 - 著作権処理
 - 契約情報・書誌情報記述、分類
 - 収集・再収集条件設定
 - 収集指示
- 品質検証
- トリミング、個体登録
- 書誌登録
- ◆ 再収集
 - 自動再収集、更新検知
 - 強制再収集
 - 品質検証
 - トリミング、個体登録

2002/12/14

74

収集対象の発見、調査

- ◆ 発見、リストアップ
 - ネットサーフィン+既存のリンク集活用
 - 収集対象かどうかの判断
- ◆ 調査
 - 技術的に収集可能かどうか判断
 - 構成ファイルのドメイン・ディレクトリ調査
 - 著作権処理範囲の決定
 - 書誌情報調査、著作権許諾先調査

2002/12/14

75

処理すべき著作権

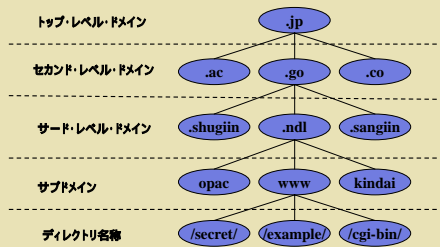
- ◆ 複製権
 - 電子データをダウンロード、複製、蓄積
- ◆ 同一性保持権
 - リンクの張替え
 - 文字コード変換
 - 長期保存
- ◆ 公衆送信権
 - 当館ホームページから利用・提供

2002/12/14

76

著作権処理範囲-URLの階層構造-

(例) <http://www.ndl.go.jp/example/>

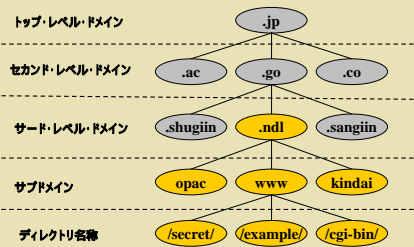


2002/12/14

77

著作権処理範囲-許諾範囲指定-

(例) 許諾する範囲を「ndl.go.jp」と指定

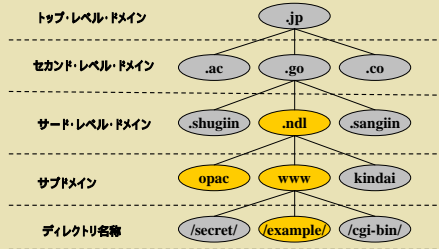


2002/12/14

78

著作権処理範囲-除外範囲指定-

(例)「kindai.ndl.go.jp」「/secret/」「/cgi-bin/」



2002/12/14

79

著作権処理範囲-電子雑誌の場合-

◆ 日本全国書誌

- URL:

http://www.ndl.go.jp/jp/publication/jnbwl/jnb_top.html

- 本文: www.ndl.go.jp/jp/publication/jnbwl/

- 画像: www.ndl.go.jp/images/
許諾は「ndl.go.jp以下」で

2002/12/14

80

許諾契約文書の内容

◆ 著作権

- 複製権、同一性保持権、公衆送信権
- 二次利用の防止

◆ 契約内容

- 収集・保存に関わるもの
- 提供に関わるもの

2002/12/14

81

許諾回答用フォーム

ウェブ サイト	ウェブサイトの名称	国会電子フォー
	URL	http://www.ndl.go.jp
	機関名	国会図書館
	最初に公開した日	1998年4月1日
許諾する 範囲 について 希望する 除外範囲 を指定する 場合は 必ず 記載する	<input checked="" type="checkbox"/> 収集・保存を許諾する	許諾する範囲: ndl.go.jp 以下、kindo.go.jp以下
	<input type="checkbox"/> 以下の条件により収集・保存を許諾する	以下: 02/15
	<input type="checkbox"/> ndl.go.jp	許諾する範囲: kindo.go.jp
	<input type="checkbox"/> ndl.go.jp	許諾する範囲: kindo.go.jp
	<input type="checkbox"/> ndl.go.jp	許諾する範囲: /secret/
	<input type="checkbox"/> ndl.go.jp	許諾する範囲: /cgi-bin/
	<input type="checkbox"/> ndl.go.jp	許諾を除外する 範囲:
	<input type="checkbox"/> ndl.go.jp	許諾を除外する 範囲:
	<input type="checkbox"/> ndl.go.jp	許諾を除外する 範囲:
	<input type="checkbox"/> ndl.go.jp	許諾を除外する 範囲:
他の権利 者に 対する 許諾 の範囲 について	<input checked="" type="checkbox"/> インターネットを通じての提供及び国会図書館構内情報通信網を通じての提供 権への利用を許諾する	条件: 1年) 1月) 1日以後なら可
	<input checked="" type="checkbox"/> 収集日から起算して: 90 日以後なら可	条件: 1年) 1月) 1日以後なら可
	<input type="checkbox"/> 国会図書館構内でのみ提供可能とした提供(国会図書館を目的とした提供)	条件: 収集日から起算して: 1日以後なら可
	<input type="checkbox"/> 提供は許諾しない	提供は許諾しない

2002/12/14

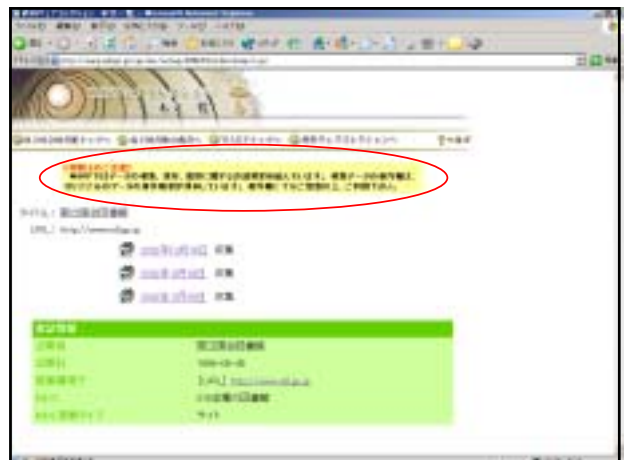
82

利用・提供条件

- インターネット上で即利用提供可能
- 一定日以後にインターネット上で利用提供可能
- 収集日より起算して一定期間経過以降にインターネット上で利用提供可能
- 国立国会図書館の館内でのみ提供可能(通常の図書や雑誌と同様)
- 提供不可

2002/12/14

83



問い合わせ対応

- ◆ 電子雑誌の定義
- ◆ 著作権
- ◆ ドメイン・ディレクトリ範囲
- ◆ 納本制度
- ◆ 実験終了後の取扱い

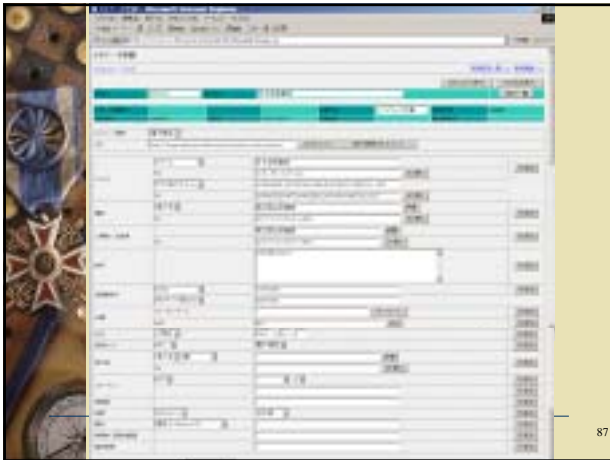
2002/12/14

85



2002/12/14

86



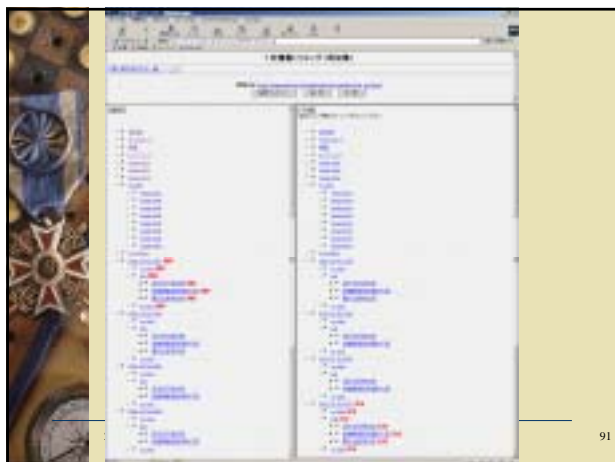
87



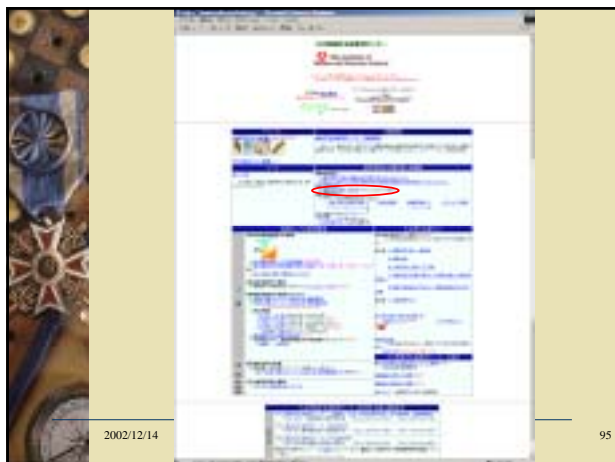
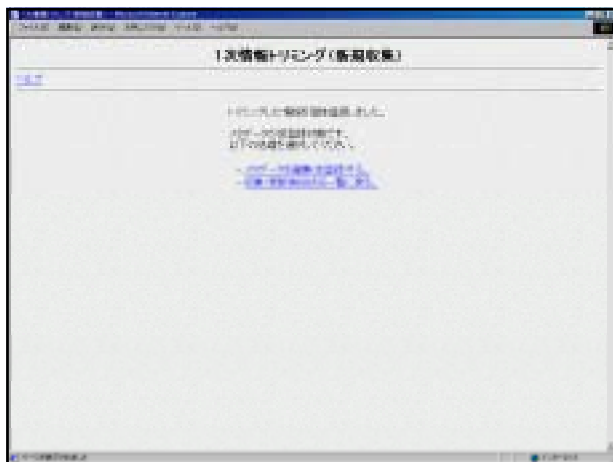
2002/12/14

88





91



2002/12/14

95

実験実施途中の課題整理

- ◆ 著作権許諾処理、契約に関わる課題
 - 手間がかかる / 著作権許諾手続きの不在と著作権者特定の困難 / 著作権者過多の問題 / 著作権者以外の利害関係者の存在 / 許諾契約内容に対する理解不足 / ウェブの頻更新性に応じた許諾契約内容変更のための作業コスト / 著作権非許諾の構成ファイルに起因する文書の不完全性の問題 / ウェブの膨大性と契約処理コスト 等
- ◆ 収集対象に関わる課題
 - 収集対象の定義に関する課題 / 収集対象の把握に関する課題 / 収集対象の範囲指定に関する問題

2002/12/14

96

実験実施途中の課題整理

- ◆ 収集処理に関わる課題
 - 自動収集時のトラブル / 収集過多の問題 / ベスト・エフォート性の問題 / 収集処理中の更新に関する問題
- ◆ 収集済電子情報の管理に関わる課題
 - 品質管理のためのガイドライン・検証方法の不在 / 収集対象文書の欠陥 / ブラウザによって再生状態の異なる場合の品質管理 / リンクの人為的切断に起因する文字化け等の問題 / サーバーのヘッダー情報の問題 / データの同一性保持について

2002/12/14

97

実験実施途中の課題整理

- ◆ 収集できないウェブ情報への対応
 - 動的生成文書 (CGI等)、登録・有料サイト、データベース等
- ◆ 組織化・利用提供上の課題
 - 時系列蓄積された膨大な文書の検索方法に関する課題

2002/12/14

98

WARPの目指すべき方向性

- ◆ 目標
 - 納本制度審議会での検討に役立てる
 - 業務モデルの確立 / コストの把握
 - 法・制度のあり方 / 基準の策定
 - 作成者・利用者に対する理解促進
- ◆ 理念
 - 文化資産としてのウェブのあるべき姿
 - 国として取り組む必要性
 - 国民的合意の形成

2002/12/14

99

もっと詳しく知りたい方へ

- ◆ 廣瀬信己「北欧諸国におけるウェブ・アーカイブの現状と納本制度」
 - 『国立国会図書館月報』490号(2002.1)
- ◆ 廣瀬信己「消えゆくウェブを救え！～動き出すウェブ・アーカイブ～」
 - 日本データベース協会発行『データベースNo.21』(2002)
- ◆ 河合美穂「Domain.uk -英国のウェブ・アーカイブ～」
 - 国立国会図書館発行『カレント・アウェアネス』No.273
2002.09.20発行

2002/12/14

100

ご静聴ありがとうございました

- ◆ WARP
 - <http://warp.ndl.go.jp>
- ◆ Dnavi
 - <http://dnavi.ndl.go.jp>

nhirose@ndl.go.jp

2002/12/14

101