

平成25年度
学術情報システム総合WS
中間報告

テーマ: デジタル化された資料の状況調査と組織化

グループ1:

名古屋大学 小島由香

日本原子力研究開発機構 長屋 俊

大阪大学 藤江雄太郎

本日の発表内容

1. 課題設定とアプローチ
2. スケジュールと作業分担
3. 現段階での調査結果と分析
 - A) 国内ポータル標本調査(愛知県・大学)
 - B) NACSIS-ILL謝絶ログの分析(2012年度分)
 - C) 文献/報告書の収集
4. ここまでの問題点・論点整理
5. 今後の予定について

1. 課題設定とアプローチ

デジタル化資料について問題は大きく以下の2つだと考えました。

A) そもそも存在を「知らない」

- ポータルの存在や、資料種別の存在などさまざま次元で「知らない」

B) あるはずだけど「たどり着けない」

- 探しているものがWeb上にあるという情報は持っているが、「たどり着けない」

1. 課題設定とアプローチ

A) 「知らない」について

目標	国内外の主要データソース/ポータルの見取り図の作成
背景	図書館員がそもそも「知らない」。知らないので、利用者に情報提供もできない。
調査内容	NACSIS-ILL謝絶のログ解析 — 潜在的ニーズの把握
	国内外の状況調査(基本情報と関係性を中心に) — ポータルの収録内容について知る — (8/1追加) 国内デジタル化資料の標本調査。全体像把握のあしかりに。

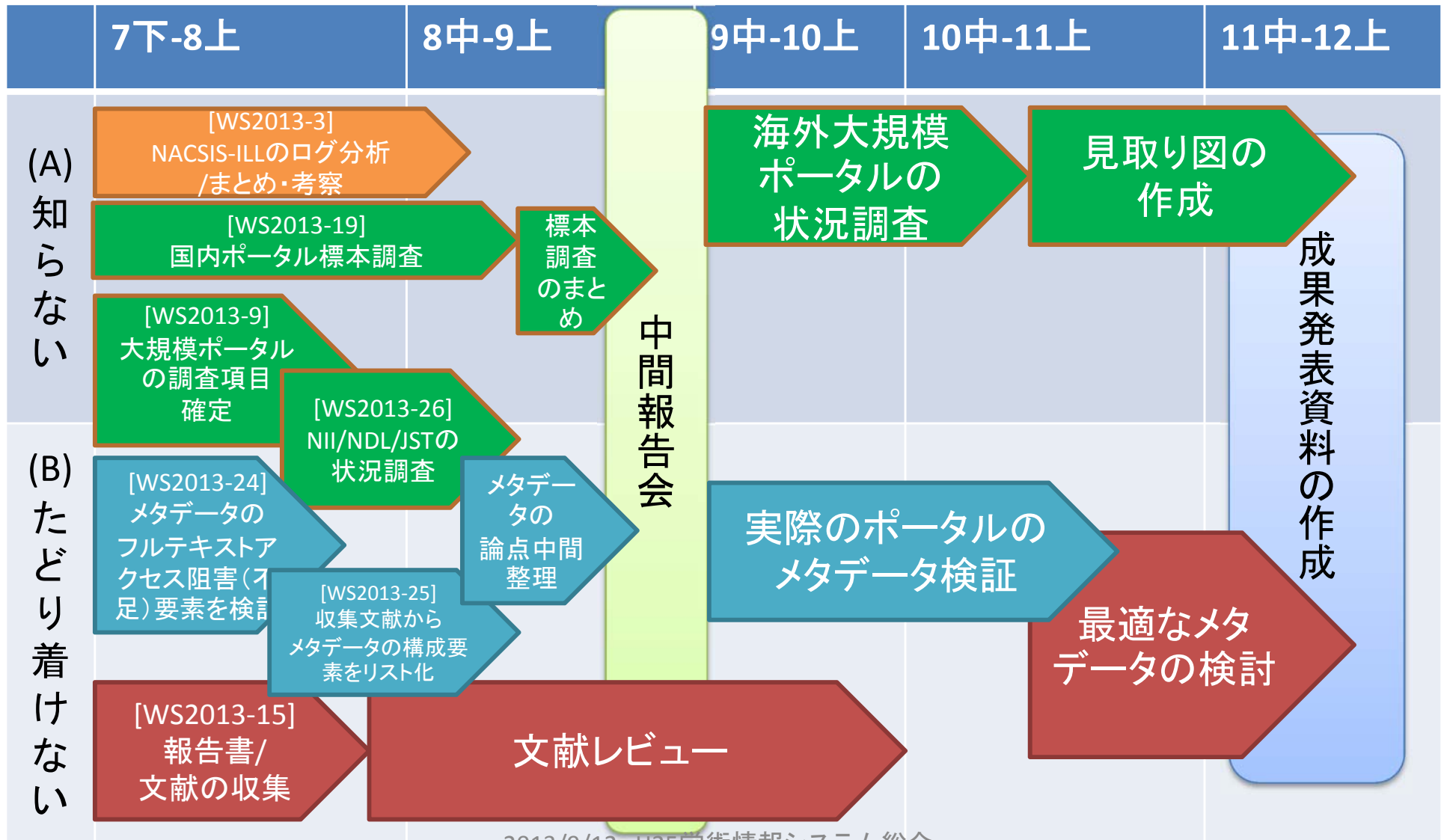
1. 課題設定とアプローチ

B) 「たどり着けない」について

目標	デジタル化資料のメタデータ/組織化についての提言の作成
背景	データソース/ポータルによってメタデータが不揃いなことや、名寄せができていないことが原因で、探しているものに「たどりつけない」
調査内容	メタデータに関する論点整理 —メタデータ形式や流通、名寄せなどさまざまな事柄があるので、まずグループで論点整理
	国内外のメタデータに関する報告書のレビュー —今後必要と思われる構成要素とメタデータ組織化の手法の抽出を目指す

2. スケジュールと作業分担

8/1改訂版



2. スケジュールと作業分担

	※★の人がリーダー	小島	長屋	藤江
(A)知らない	調査項目の確定	●	●	★
	－国内の状況調査	●	●	★
	－海外の状況調査			★
	NACSIS-ILLのログ分析			★
(B)たどり着けない	メタデータの論点整理	●	★	●
	－実際のポータル メタデータ検証	●	★	
	報告書/文献の収集	★	●	
	－文献レビュー	★	●	

3.現段階での調査結果と分析

A)国内ポータル標本調査(愛知県・大学)

【調査の目的】

- 実際に1つ1つのデータを見ることで、気づいてなかった「デジタル化資料」がないか探る
- 国内のデジタル化資料の悉皆調査の実行可否を探る
- 悉皆調査の際の調査項目の検討材料にする

【対象】

愛知県の国公私[○]の4年制大学51校

【方法】

- 大学図書館ホームページのナビゲーションから探す
- 大学トップページの目次やサイトマップからリンクからとべる範囲で探す
- 次ページの調査項目を設定したEXCELを用いてチェックした
- 3人で担当大学を割り振って分担調査

3.現段階での調査結果と分析

A)国内ポータル標本調査(愛知県・大学)

【調査項目】

- 提供コンテンツの分類
- 提供コンテンツの量
- ポータル種別
- 大学図書館HPからのナビゲート有無
- OPACへのメタデータ収録有無
- リゾルバへのメタデータ収録有無
- 統合検索/ディスカバリへのメタデータ収録有無
- リポジトリへのデータ登録
- Googleから全文検索可能か
- Googleから個々のコンテンツ単位でヒットするか

3. 現段階での調査結果と分析

A) 国内ポータル標本調査(愛知県・大学)

【分析対象】

標本調査完了後、カテゴライズを行い、下記2パターンで分析

① 学術コンテンツのみ

紀要・研究報告、研究成果(芸術)、
貴重書・古文書等デジタル化、電子書籍

② 学術コンテンツ+広報誌

①に広報誌・ニュースレター、
図書館報を追加

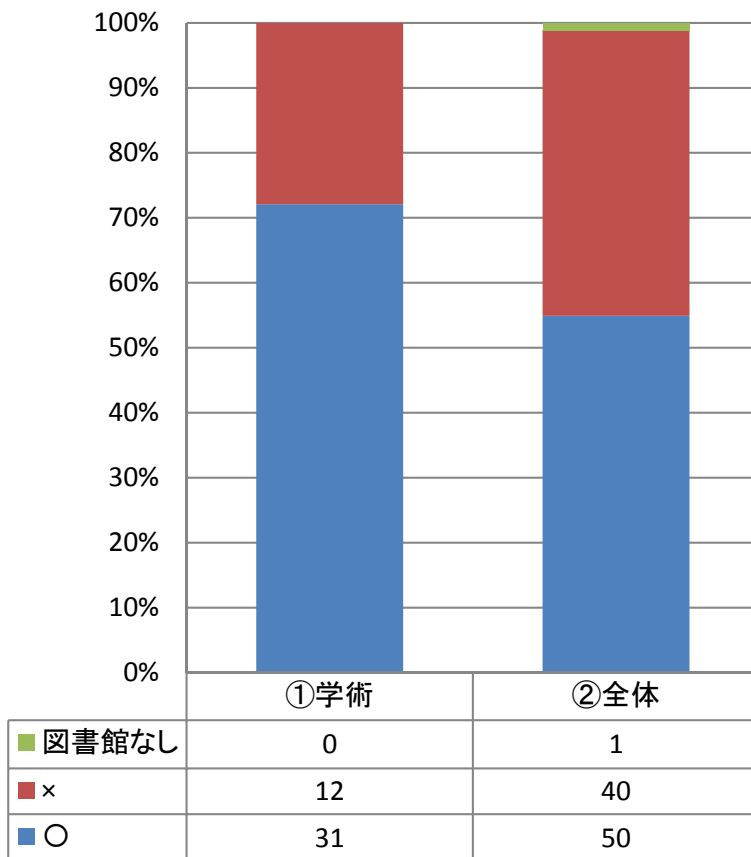
種別	件数
紀要・研究報告	23
研究成果(芸術)	4
貴重書・古文書 等デジタル化	15
電子書籍	1
広報誌・ニュー スレター	33
図書館報	15
総計	91

3.現段階での調査結果と分析

A)国内ポータル標本調査(愛知県・大学)

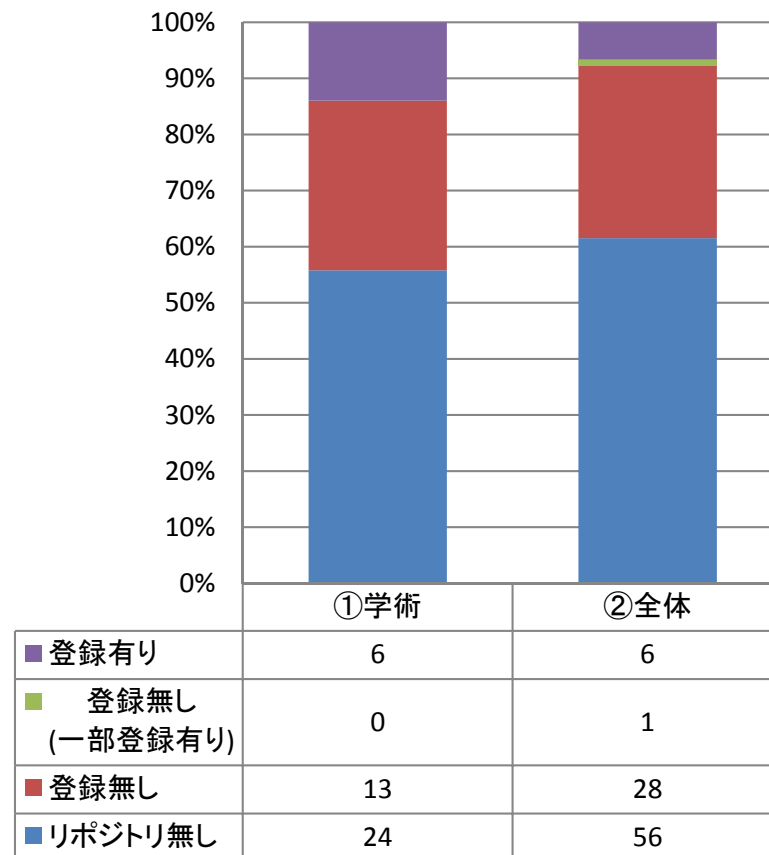
【分析①】

◆図書館HPからのナビゲート



学術コンテンツの方が大学図書館HPからのナビゲーションがなされている

◆リポジトリへの登録有無



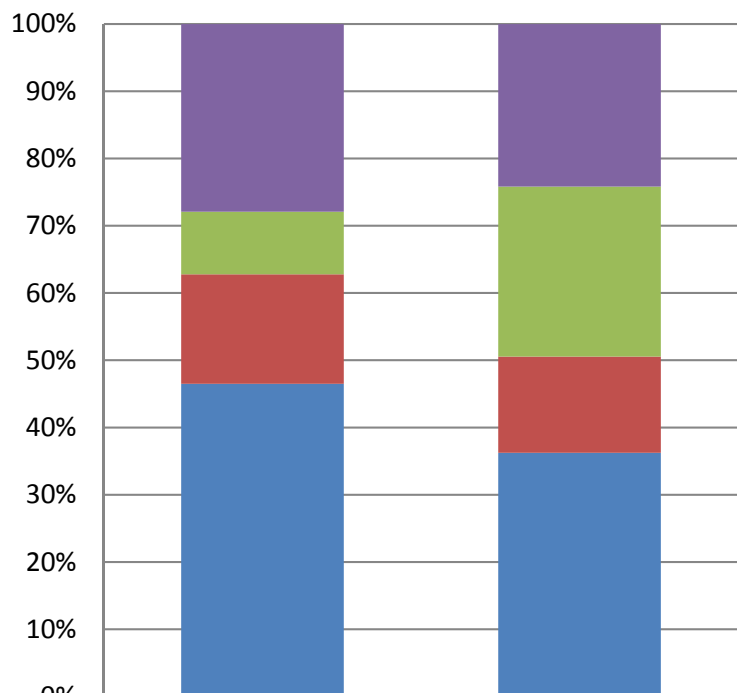
リポジトリがまだ無い大学も多く、有意な結果は得られなかった

3.現段階での調査結果と分析

A)国内ポータル標本調査(愛知県・大学)

【分析②】

◆Googleからヒットするか



	①学術	②全体
どちらも×	12	22
全文のみヒット	4	23
個々のみヒット	7	13
個々も全文もヒット	20	33

②に比べて①は「全文のみヒット」の割合が小さい。学術コンテンツは広報誌に比べて書誌情報を作成してあることが分かる。

学術コンテンツの中には、仕組み上Googleの検索でうまくヒットしないものがある。これらはページの作り方を変えたり、外部へのデータ提供を行わないと発見性は低いままだと考えられる

◆OPAC/リゾルバ/DSへの登録

全体を見ても、ほぼデジタル化資料の登録は皆無であった。

3. 現段階での調査結果と分析

A) 国内ポータル標本調査(愛知県・大学)

【全体考察】

目的に照らし合わせて

- 気づいてなかった「デジタル化資料」→成果物としての動画や芸術作品など。
- 国内のデジタル化資料の悉皆調査の実行可否を探る
→1機関あたり15分。大学だけでも800弱あるので、3人で悉皆調査するのは困難
- 悉皆調査の際の調査項目の検討材料にする
→今回の調査項目に加えて、外部へのデータ提供有無とその手段や、NII/NDL/JSTなどのポータルとの連携関係もみるとよさそう

その他気づいたこと

- デジタル化資料のデータ作成方法による可視性の低下の問題
→これは各機関でデータの作り方や外部提供について改善してもらうしかなく、外からはどうしようもない
- 調査対象の「デジタル化資料」をきちんと定義する必要性

3. 現段階での調査結果と分析

B) NACSIS-ILL謝絶ログの分析(2012年度分)

【調査の目的】

ニーズはあるが発見性の低いコンテンツをあぶりだす

【対象】

最終更新日が2012年度で状態がCANCELのNC-ILLレコード

【方法】

下記の手順でデータを絞った

1. URLが入っているSEND CMNTを抽出
2. それらを1つずつチェックして、デジタル化資料へ案内して謝絶しているコメントを絞った

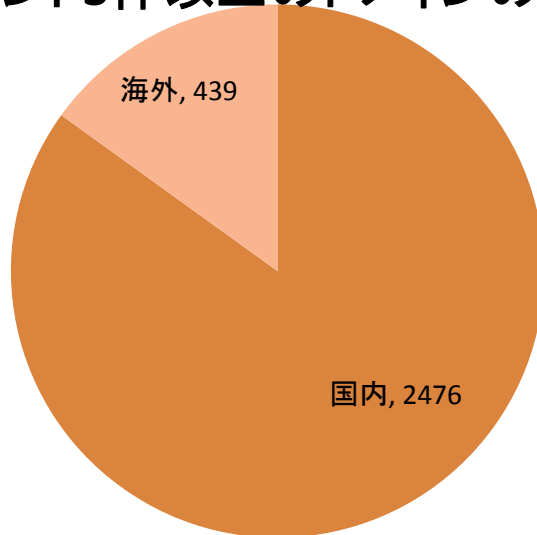
全体の分析は藤江・長屋、NII/NDL/JSTドメインのコンテンツの詳細な分析は長屋が行った。

3.現段階での調査結果と分析

B) NACSIS-ILL謝絶ログの分析(2012年度分)

【全体分析①】

カウント3件以上のドメインの分布



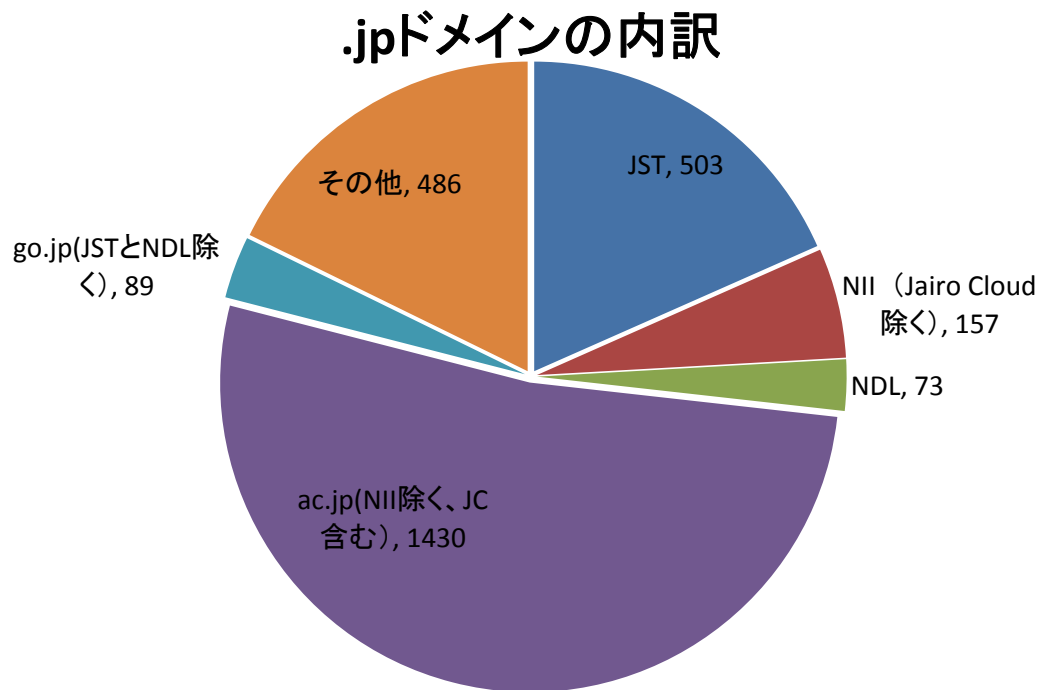
●ドメイン単位のカウントが累積3件以上のものをみてみたところ、大半が国内のコンテンツであった。(.jp以外はリンク先を1つずつチェックした)

●海外コンテンツはほとんどがコンテンツ提供元の出版社・学会のページへの案内であり、機関リポジトリも含めて第3者の運営するデジタル資料提供ポータルへの案内は少ないようであった。

3.現段階での調査結果と分析

B) NACSIS-ILL謝絶ログの分析(2012年度分)

【全体分析②】



「フリーのデジタル資料が存在している」、とILLの謝絶コメントで返すもの(※1)のうち、ドメインのURLが.jpで終わるもの。

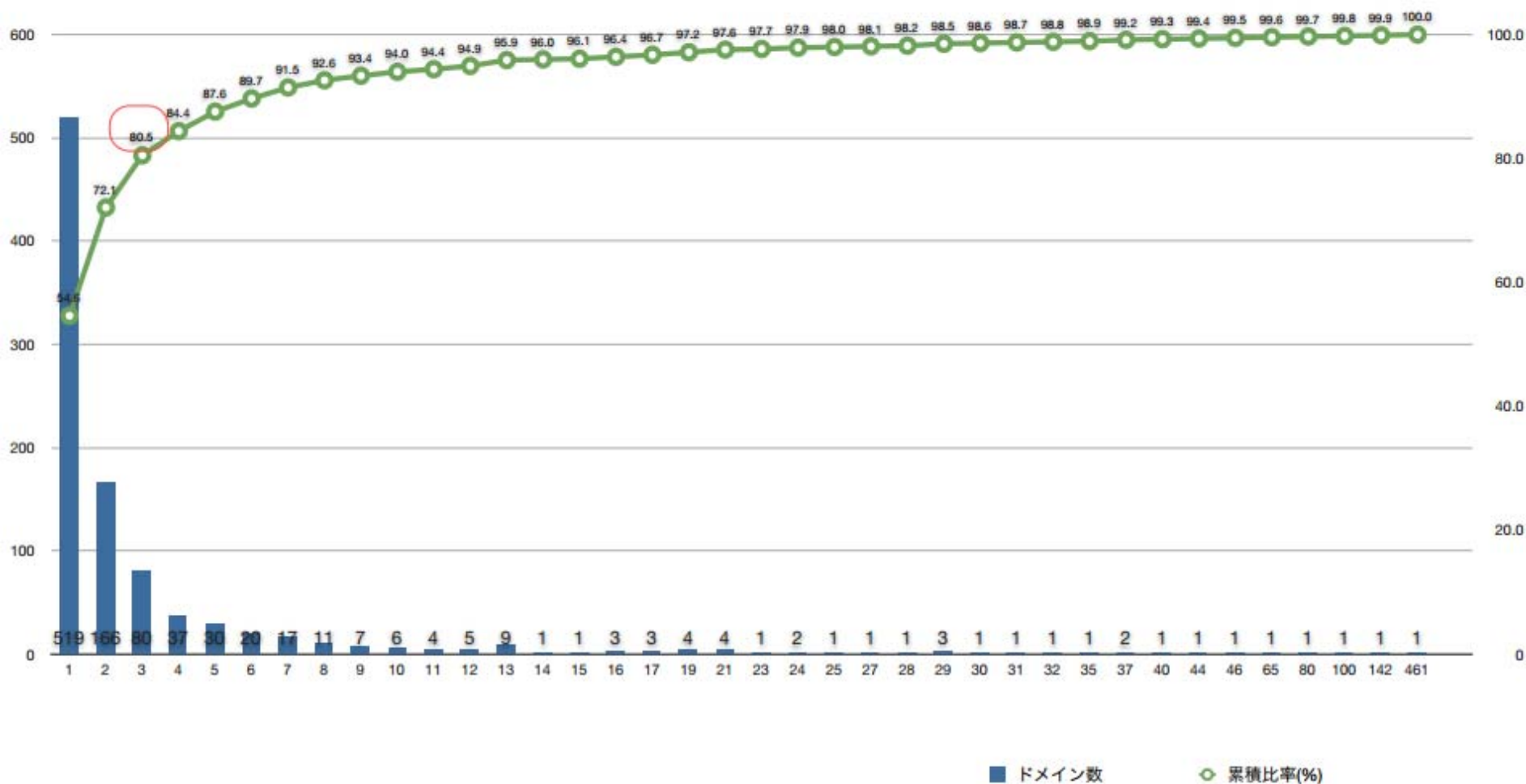
「※1」の全体が3766件あり、そのうち.jpで終わるものは2738件。
その中での分布は左記のグラフのようになった。

.jpで終わるもののうち、およそ半分がac.jpドメインで終わるものであった。機関リポジトリを中心に大学の提供しているフリーコンテンツへの案内が多いようである。また、.jpドメインのうち、1/4程度はJST・NII・NDL提供のコンテンツであった。

3.現段階での調査結果と分析

B) NACSIS-ILL謝絶ログの分析(2012年度分)

【全体分析③-1】



3.現段階での調査結果と分析

B) NACSIS-ILL謝絶ログの分析(2012年度分)

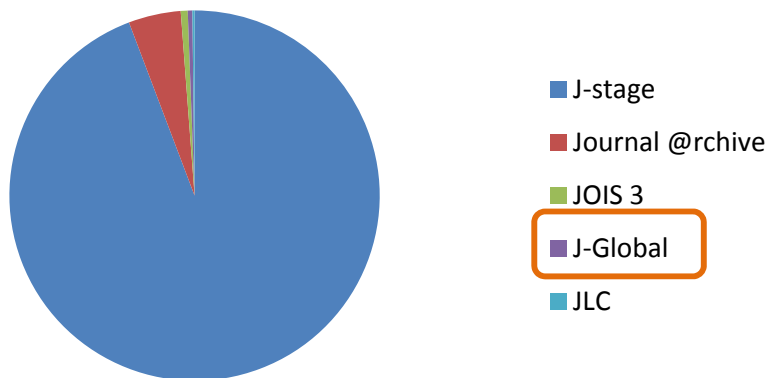
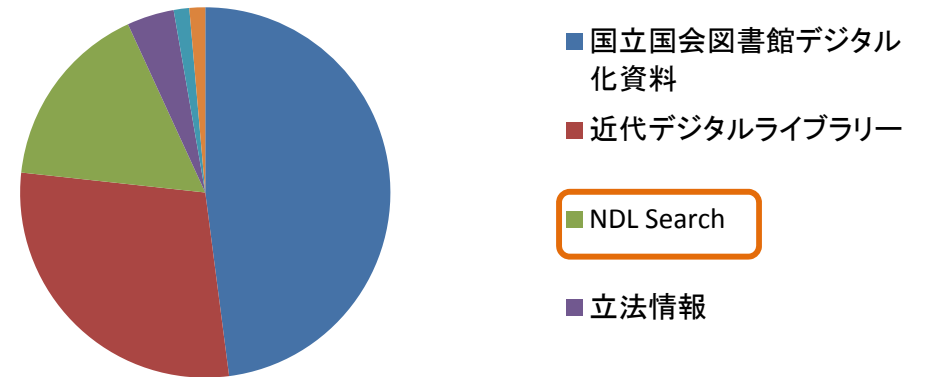
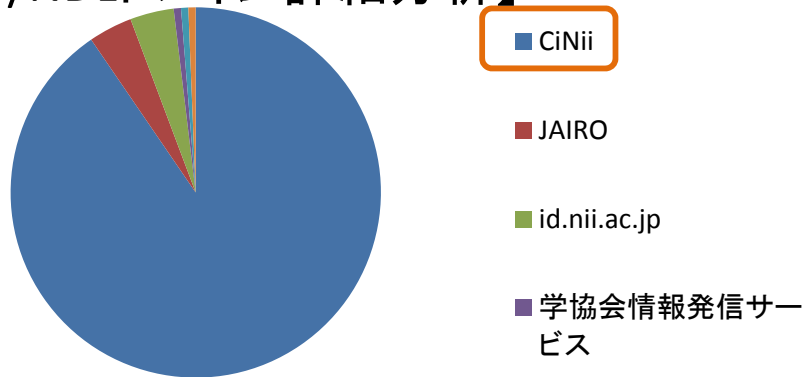
【全体分析③-2】

- 100回 ~ 461回も参照されているドメイン(NII,JST,NDL等)
 - デジタル化資料を所有する大手ドメインとそこへの需要が存在
 - 1 ~ 3回しか参照されていないドメインが全体の8割
 - デジタル化資料はウェブ上のあちこちに存在
- ⇒ デジタル化資料を探すにはポータルと検索エンジンの併用が必要、と推測できる

3. 現段階での調査結果と分析

B) NACSIS-ILL謝絶ログの分析(2012年度分)

【NII/JST/NDLドメイン詳細分析】



全てのコンテンツではないが、それぞれのドメインにおいて、

- NII – GeNii
- JST – J-Global
- NDL – NDL Search

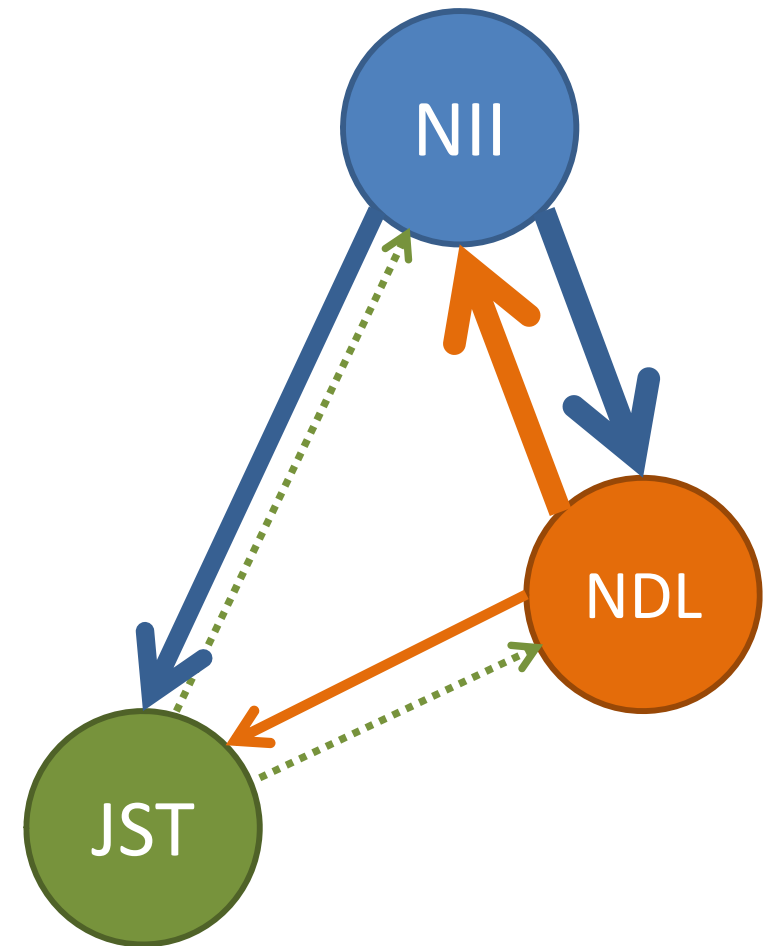
で検索可能

3. 現段階での調査結果と分析

B) NACSIS-ILL謝絶ログの分析(2012年度分)

【補足:ポータル間の関係】

- CiNii
 - JST : J-Stage 43万件
 - NDL : 雑誌索引記事 1037万件
- NDL Search
 - CiNii Article 400万件
 - Books 1000万件
 - J-Stage 15万件
- J-Global
 - なし(リンク関係のみ)



⇒ 全てのコンテンツを対象に検索するにはそれぞれで検索する必要がある

3.現段階での調査結果と分析

B) NACSIS-ILL謝絶ログの分析(2012年度分)

【全体考察】

- 全体分析①: 海外コンテンツについて件数が少ないのはニーズの問題なのか、図書館員が知らないだけなのか。今後WSで海外も取り扱うならば、もう少し詳細に見る必要あり。
- 全体分析③: ウェブのいたるところにデジタル化資料が散らばっている。
 - 悉皆調査の必要性。WSで対象とするのはどこか? ILL謝絶データも元になる?
 - 関係機関が多い。メタデータを整備してもらうのは困難。何かやり方はある?
- 全体分析③+詳細分析: 大手ドメイン(=サービス)でかなりのコンテンツを持っていて需要がある。
 - 少なくとも大手はしっかりメタデータを発信して欲しい。WSを通じて調査、検証、提案くらいはできる?
 - (どこでもリゾルバのような横断で調べる仕組みがあると便利)
- JSTは可視性が低いように考えられる。ILLの謝絶トップ。なおかつ、他のポータルから検索対象とされていない。(保有コンテンツやポータルの包含関係については詳細未確認)。
- 謝絶コンテンツについて、国内ポータル標本調査のように項目を整理してまとめると、可視性の低さが何に起因するのか見えてくる可能性。

3. 現段階での調査結果と分析

C) 文献/報告書の収集

【調査の目的】

- ・メタデータ形式や流通、名寄せなどさまざまな事柄があるので、論点整理の材料に
- ・今後必要と思われる構成要素とメタデータ組織化の手法の抽出を目指す

【対象雑誌】

情報の技術と科学	Vol.24(1) (1981)～Vol.56(3)(2013) ※2009年以前はキーワードで検索	論文数: 46
情報管理	Vol.49(1)(1999) ～Vol. 63(7) (2013)	論文数: 56
図書館雑誌	103(1)(2009.01)～107(1)(2013.01)	論文数: 21
日赤図書館雑誌	(おまけ)	論文数: 3

【方法】

- 雑誌を目次からブラウジングして関連のありそうな文献を探して、Backlogにて共有
- 小島が収集を担当

3.現段階での調査結果と分析

C) 文献/報告書の収集

【分析】

論文をテーマごとに分けるとこのような分布になった。

論文のテーマ	資料の分類	論文数
検索エンジンの技術	I-A	3
Googleの技術	I-A-(1)	5
リテラシー教育	I-B	3
ディスカバリーサービス	I-C	5
デジタル資料:総論	II	6
デジタルコレクション	II-A-(1)	1
研究データ	II-B-(1)	3
学内刊行物	II-B-(2)	1
講義資料	II-B-(4)	1
貴重資料	II-B-(5)	7
XML	II-C-1-(4)	2
サーバ	II-D	1
リポジトリ	II-E-1-(1)	3
デジタル資料のメタデータ	II-E-2	5
DOI	II-E-3-(2)	1

論文のテーマ	資料の分類	論文数
電子図書館:総論	III	10
目録規則	III-A-1-(4)	3
書誌ユーティリティー	III-A-1-(6)	2
リゾルバ	III-A-2-(1)	4
ERMS	III-A-2-(2)	1
大規模なデジタル化	III-B-(1)	4
デジタル化プロジェクト	III-B-(2)	11
アーカイブ	III-B-(3)	11
権利関係	III-B-(4)	2
ポータル	III-C	3
ポータル技術	III-C-2	6
API	III-C-2-(2)	1
FRBR	III-C-2-(4)	1
著者名典拠	III-C-2-(5)	7
メタデータ	III-C-2-(7)	13
	総計	126

4. ここまでの問題点・論点整理

【問題点】

- 使っている用語について3人で認識にズレがあるので再定義の必要
 - Ex.「調査対象の『デジタル化資料』」、「メタデータ」、「ポータル」
 - 手がかりを探す意味合いも込めて広範な調査計画を立てた。が、しかし…
 - それぞれの調査の焦点がぼやけがち
 - 作業量が多くて、こなすことで精いっぱいになってしまっている
- ⇒論点の絞り込みを行う必要性。

4. ここまでの問題点・論点整理

【論点整理】

- 幸いこれまでの調査でさまざまな論点を認識することはできた。
- 別添「【1)チームfujimaya】テーマの絞り込みのための資料.docx」を用いて絞り込みを行う。

5. 今後の予定について

第2回WS時に4.の論点整理を経たテーマの絞り込みを行う



それに基づいて12月までの計画を設定する