

第6回東京大学学術資産アーカイブ化推進室主催セミナー
デジタル資料の保存管理の実際

東京大学での取り組み
史料編纂所

山田太造

東京大学

史料編纂所/ 地震火山史料連携研究機構

自己紹介

■山田 太造(やまだ たいぞう)

■専門: データ工学

- 博士(情報学)
- 問い合わせ処理, 分散処理, 機械学習, メタデータ
- 歴史情報学(, 人文情報学)

■経歴

- 国立情報学研究所 特任研究員 (2005-2007)
- 東京大学史料編纂所 特任助教 (2007-2010)
- 人間文化研究機構本部 特任助教 (2010-2013)
- 東京大学史料編纂所 助教 (2013-2019)
- 東京大学地震火山史料連携研究機構 (2017-)
- 東京大学史料編纂所 准教授 (2020-)

■e-mail: t_yamada@hi.u-tokyo.ac.jp

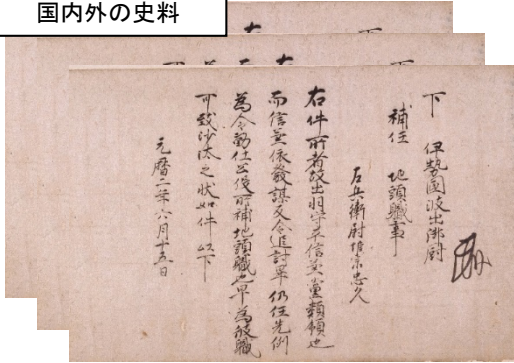
史料編纂所における研究活動

史料採訪

編纂

出版

国内外の史料



収集・調査

メタデータ

- 文書名: 源頼朝下文
- 所蔵先: 東京大学史料編纂所
- 和暦: 元暦2年6月15日
- ...
- デジタル画像
- JPEG 9184x7239 8-bit sRGB

翻刻

索引

再編成

網文

下 伊勢国波出御厨
補任 地頭職事
左兵衛尉惟宗忠久
右件所者故出羽守平信兼
...

オブジェクト識別・同定

- 史料
- 人名
- 地名
- ...

平信兼黨
類ノ所領

額朝忠久
ヲ伊勢波出
頭御厨地
ニ補任

下 伊勢国波出御厨
補任 地頭職事、
(異筆) 左兵衛尉惟宗忠久
(頼朝) (花押)

右件所者故出羽守平信兼黨類領也、而信兼依發謀反、令追討畢、仍任先例、爲令勤仕公役、所補地頭職也、早爲彼職、可致沙汰之狀如件、以下、

例:『大日本古文書』

- 史料集
- DB公開

史料集の出版



2つの史料集

- 編年史料
 - 事件が起きた時間の流れに沿って関連史料を集めたもの
 - 『大日本史料』, 『大日本古文書』(編年文書)など
- 編纂史料
 - 史料群の構成と内容を精密に復元
 - 『大日本古文書』(家わけ文書), 『大日本古記録』など

1901(明治34年)から, 第二次世界大戦末期と終戦後の一時期を除き, 出版事業を継続. 年間十数冊を刊行し, 総点数は約1100冊以上.

明治34年刊行の
『大日本史料』(第六編・第十二編)
『大日本古文書』(編年文書)

<http://www.hi.u-tokyo.ac.jp/about/mission>

史料編纂所のデータベース

https://wwwap.hi.u-tokyo.ac.jp/ships/shipscontroller

DB数

- 所内: 39
- 所外: 31
 - ・整備中, 公開するにはふさわしくないと判断したものは所外公開せず

種類

- 目録
- 本文
- 画像・図像・写真
- 知識
 - ・人名
 - ・地名
 - ・和暦
 - ・文字

範囲

- 日本前近代
- 奈良～維新

史料集編纂において必要な情報をデータベース化

- 画像史料解析センターの成果に基づくものも含む

データベース検索

ホーム 所員用ログイン English ヘルプ

キーワード (例: 島津家文書) 横断検索

史料の所在

所蔵史料目録データベース (HI-CAT)
Hi-CAT Plus
日本古文書ユニオンカタログ
正倉院文書マルチ支援データベース
Catalogue Database of the Batavia's Uitgaand Briefboek, 1621-1792

できごとを主題に

大日本史料総合データベース
編年史料カード (古代関係) データベース
近世編年データベース
維新史料綱要データベース
近世史編纂支援データベース (摺物データベース) (錦絵データベース)
編年史料 (古代) 編纂支援資源化データベース MIDOH

画像をさぐる

歴史絵引データベース
肖像情報データベース
史料編纂所所蔵肖像画模本データベース
史料編纂所所蔵荘園絵図模本データベース
金石文拓本史料データベース
摺物データベース
錦絵データベース
古写真データベース

全文の検索

古記録フルテキストデータベース
古文書フルテキストデータベース
奈良時代古文書フルテキストデータベース
平安遺文フルテキストデータベース
鎌倉遺文フルテキストデータベース

人を主題に

(大日本史料総合データベース)
中世記録人名索引データベース
花押データベース
志形見データベース
大日本史料7編人名カードデータベース (肖像情報データベース)
(史料編纂所所蔵肖像画模本データベース)

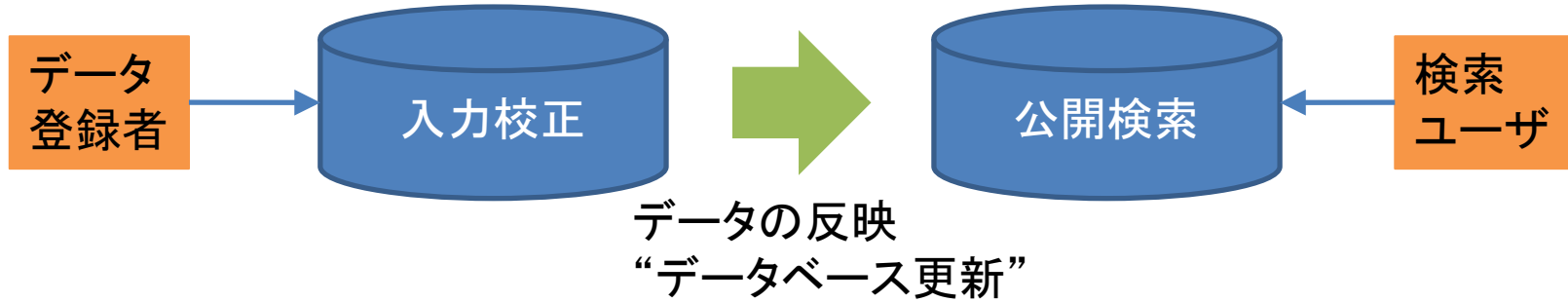
辞典・字典

応答型翻訳支援システム(ダウンロードページへ)
欧文日本古代史料解題辞典データベース(Dictionary of Sources of Classical Japan)(Dictionnaire des sources du Japon)
電子くずし字字典データベース
史的文学データベース連携検索システム

登録件数: 約725万レコード, 約2000万画像, 約9.5億文字
アクセス数: データベース 約440万件/年, 画像 約1,056万件/年

データ作成 → 公開

SHIPS DB



ex. 所蔵史料目録DB(Hi-CAT)

所蔵史料目録データベース (書目データ更新画面) DB選択 | 処理選択 | 簡易検索 | ログアウト

前件 次件 先頭 最終 検索結果一覧

3/1,335件

書目ID : 00010961

書名 : 三宝院伝法灌頂私記 内道場

書名ヨミ : サンホウイ

区分コード : 000 原本・古写本類 (0架)

別冊記号 :

架番 : 架 0014 番 1 枚番

原区分 :

別冊記号 :

原架番 : 架 番 枚番

著者名 : (伝)静温

原蔵者 :

原蔵者ヨミ :

原蔵者所在地 :

出版事項 : 写本 本奥書 全秀

形態1 : 1巻 37cm(卷子) 参照

撮影・複本作成 :

撮影済み :

引伸 :

大きさ :

所蔵史料目録データベース DB選択 | 項目検索 | 簡易検索 | ヘルプ

前件 次件 先頭 最終 一覧

4/1319件

《書目データ》

【書目ID】 00010961

【史料種別】 貴重書 (原本・古写本類)

【請求記号】 0014-1

【書名】 三宝院伝法灌頂私記 内道場

【著者名】 (伝)静温

【出版事項】 写本 本奥書 全秀

【形態】 1巻 37cm(卷子)

【注記】 (紙背註あり)[貞応三, 正嘉三, 文正元年の奥書あり]

解題 イメージ

《冊データ》

イメージ

史料目録(メタデータ). モノの伝来由来などはある.
画像は？

所蔵史料と所外の史料

2020年3月よりweb
公開(一部)

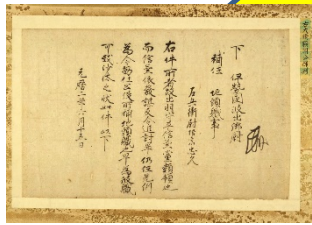
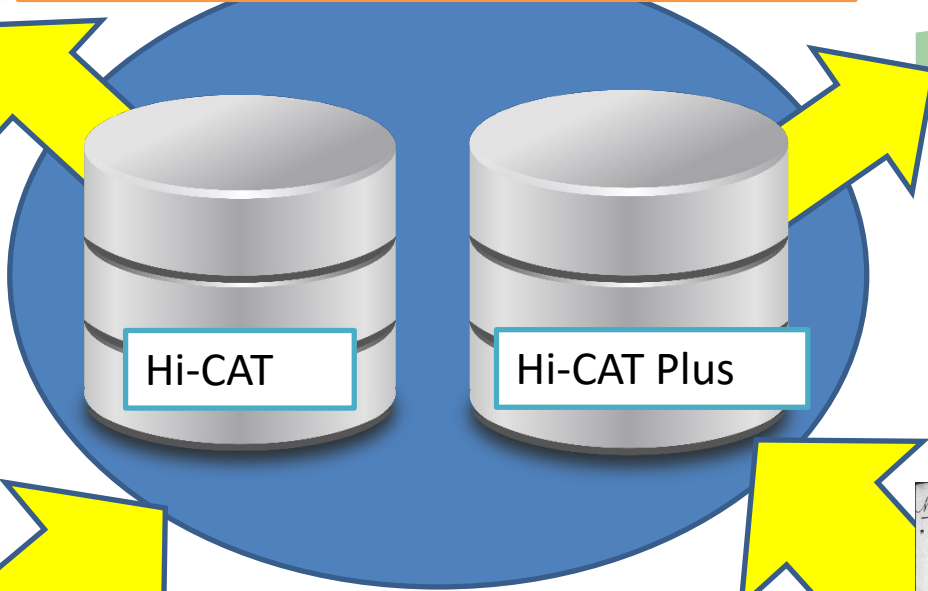
SHIPS
史料編纂所歴史情報処理システム



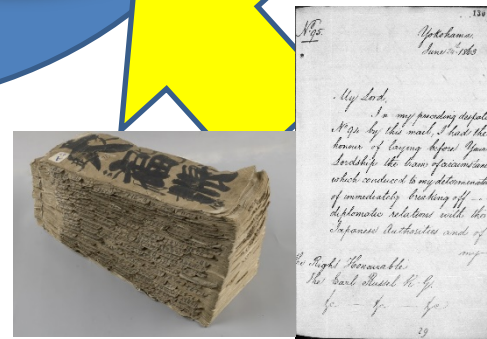
Web公開



閲覧室での公開
+ 連携機関での閲覧
例: 秋田県公文書館.
横浜開港資料館



史料編纂所所蔵史料の画像



国内・海外から収集した史料の画像

編纂と史料複製

■史料採訪 約50件(約200史料群) /1年

- 史料編纂所以外にある史料が対象
- 史料の調査及び収集(複製作成・撮影)
- 1885年より本格化, 135年間

■収集

–当初:

- 影写本(敷き写し), 謄写本(見取り写し), 模写など
- 並行してガラス乾板による撮影

–マイクロフィルム:

- 1950年より開始, 1970年ごろから本格化. ~ 2010年
- スキャニング 2008年- 2015年. 約 12,000 リール
- 国外分: 1930年頃から収集開始. デジタル化は2014年度より本格的に開始

–デジタル撮影:

- 2010年頃から

2020.2.11 都城島津邸(都城市)



2020.10.21 松尾大社(京都市)

史料採訪の流れ

規約化は 2008年開始(3年かかった)
試行は2010年, 実施は2011年から
運用後10年経過, 見直し中

史料編纂所内で規約化
デジタル採訪小委員会でフローを策定

研究者
撮影情報を図書部へ申告
含 目録・セキュリティ

史料採訪

写真室
画像変換・補正

画像整備

図書部
研究者の申請に基づく

目録整備

採訪計画・
準備

史料採訪委員会・
図書部

セキュリティ
設定

情報支援室
図書部と研究者の申請に基づく

DB登録

図書部, 情報センター, 各種プロジェクト

画像登録

図書部・情報支援室

史料の撮影・カメラの設定等
のマニュアルも用意.

史料画像デジタル化進捗管理システム

史料画像デジタル化進捗管理システム

ログアウト ヘルプ

検索結果一覧画面 > 画像管理情報参照画面

【画像管理情報参照画面】

< 前へ 次へ >

画像管理情報 (本所所蔵史料)

印刷表示

進捗確認

データコピー

画像管理情報編集

No.	項目	記入・選択項目
1	B D 番号	BD2099-YYYYYY
2	B D 番号申請者	図書館
3	画像管理情報記録者	図書館
4	画像管理情報記録年月日	2019-04-03
5	撮影対象	本所所蔵史料
6	撮影対象 (備考)	(本所所蔵史料) ←必須
7	史料 (群) 名	【記入例】加判之列勅方笈 ←書名もしくは史料名(冊次、巻次がある場合は付け加える)を入力
8	史料 (群) 名ヨミ	カハンノレツツメカタオボエ
9	原蔵者	東京大学史料編纂所
10	原蔵者ヨミ	トウキョウダイガクシリョウヘンサンジヨ
11	原蔵者所在地	東京都文京区本郷7丁目3-1
12	撮影代表者	図書館、[所外者撮影] ←必須
13	撮影ナンバー	堀内カラー[撮影] ←業者名必須
14	撮影年月日	2019-04-01~2019-04-02
【画像種別】		
15	撮影時データ備考	RAW
16	ファイル数	113
17	原蔵番号	備後福山阿部家史料-02-014 ←所蔵史料目録の架番号もしくは未整理番号を入力
18	フォルダ分化	有
19	分化後フォルダ名	BD2099-YY0000~BD2099-YY0098 ←※別途撮影リストを作成してください
20	目録注記事項	
21	撮影目録の有無	
22	利用条件確認先 (担当者名・所属)	

史料画像デジタル化進捗管理システム

ログアウト ヘルプ

検索結果一覧画面

新規登録

検索条件

B D 番号

撮影対象

史料 (群) 名

原蔵者

撮影代表者

現在の進捗状態

ステータス

クリア 検索

2 3 4 5 次へ > 最終へ >>

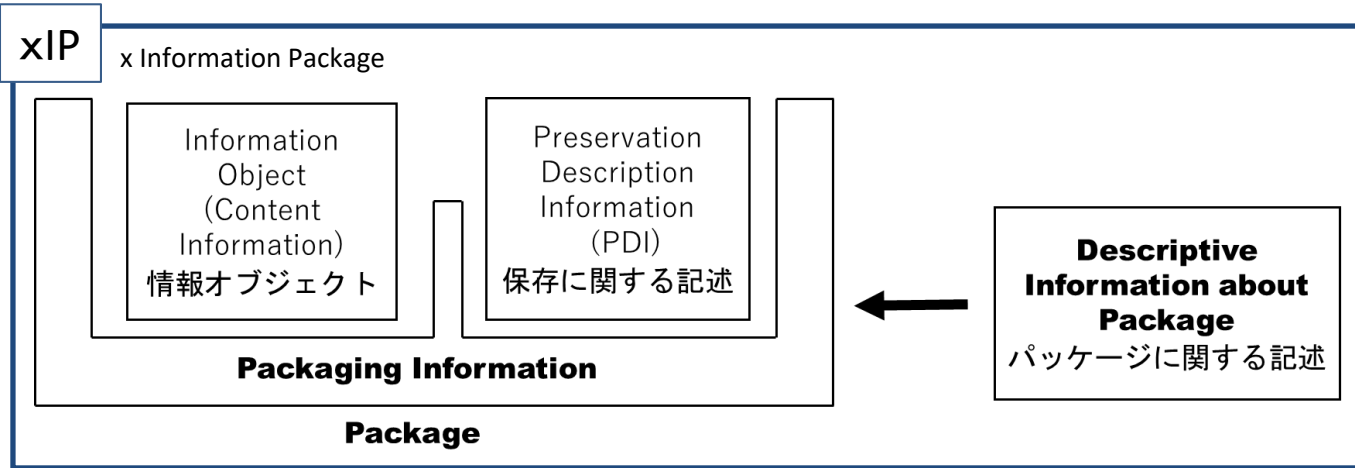
1件~20件 / 3458件中 20件 >

B D 番号	撮影対象	史料 (群) 名	原蔵者	撮影代表者	撮影年月日	進捗状態	画像管理情報
BD2099-YYYYYY	本所所蔵史料	【記入例】加判之列勅方笈 ...	東京大学史料編...	図書館、[所外...	2019-04-01...	(7) 作業中	参照 編集
BD2099-XXXXXX	本所所蔵史料	【記入例】扶桑略記 廿五 ...	東京大学史料編...	図書館	2019-04-01...	(2) 未着手	参照 編集
BD2021-007200	所外史料	醍醐寺文書 380函	醍醐寺	高橋慎一朗	2021-02	(10) 未着手	参照 編集
BD2021-007100	所外史料	小川八幡神社大般若経	小川八幡神社	遠藤基部	2021-07-23	(3) 未着手	参照 編集
BD2021-007000	本所所蔵史料	宗家史料	東京大学史料編...	荒木裕行		(6) 未着手	参照 編集
BD2021-006900	本所所蔵史料	宗家史料	東京大学史料編...	荒木裕行		(6) 未着手	参照 編集
BD2021-006800	本所所蔵史料	史談会本	東京大学史料編...	保谷 徹	2020-10-01-	(6) 未着手	参照 編集
BD2021-006700	本所所蔵史料	万里小路惟房書状	東京大学史料編...	村井	2021-07-01-	(5) 未着手	参照 編集
BD2021-006600	本所所蔵史料	称光天皇口宣案	東京大学史料編...	村井	2021-07-01-	(5) 未着手	参照 編集
BD2021-006500	本所所蔵史料	伊達政宗消息	東京大学史料編...	村井	2021-07-01-	(5) 未着手	参照 編集

もともとは独自の保存モデル
-> OAI参照モデルに準拠

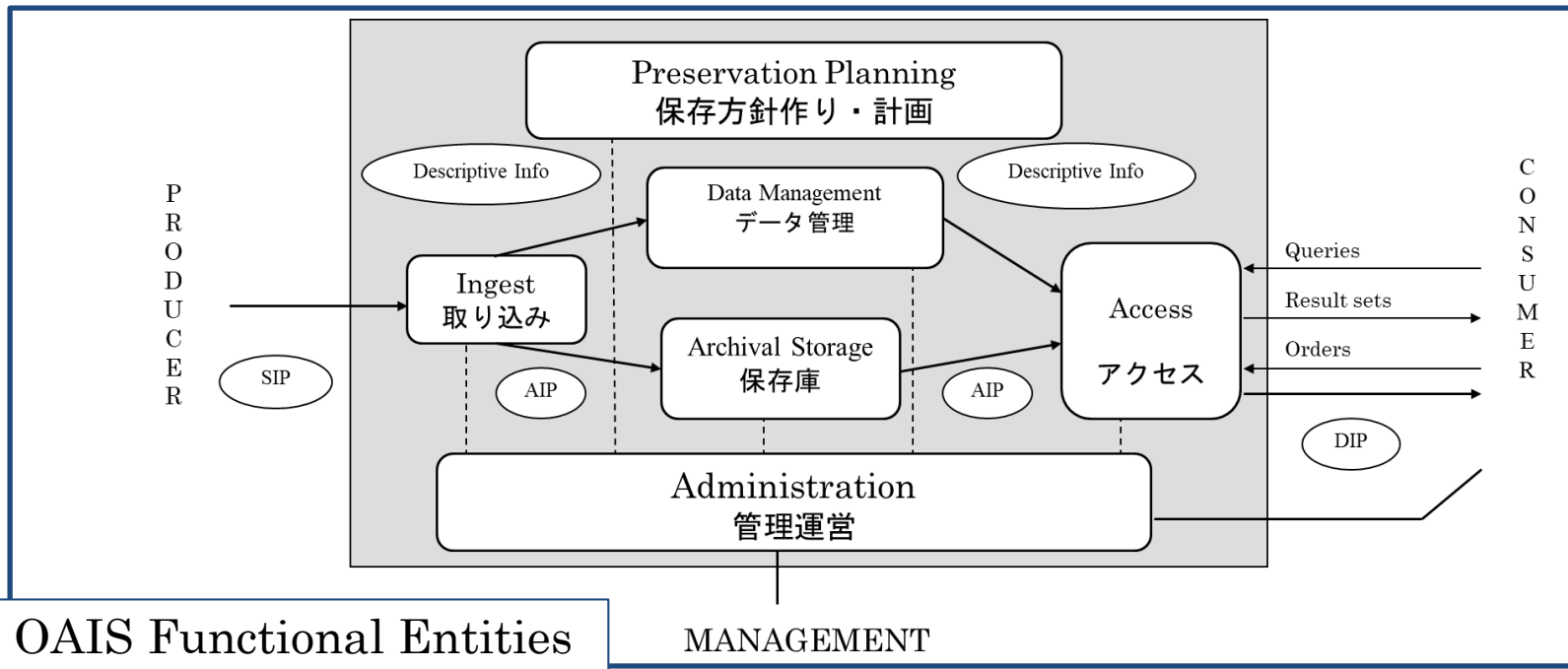
OAIS参照モデル

Reference model for an Open Archival Information System



PDI

- 来歴 (Provenance) : コンテンツの由来・伝来. 保存の処理など履歴.
- コンテキスト (Context) : コンテンツとその環境との関係. 他のオブジェクトとの関係など.
- 参照 (Reference) : コンテンツの識別子.
- 不変性 (Fixity) : オブジェクトが変更されていないことを保証する情報.
- アクセス権 (Access Rights) : コンテンツへのアクセス制限・条件.



史料所蔵者との調整

■確認方法

- 書面: e.g., 公開方法・利用について史料所蔵者と覚書を交わす
- 口頭, 撮影者判断
- (史料編纂所所蔵史料: 史料編纂所の規定による)

■書面等

- 進捗管理システムにて書類を保存(PDFなど)

22	利用条件確認先 (担当者名・所属)	<input type="text"/>
23	目録公開と利用条件の 確認方法	<input checked="" type="checkbox"/> 本所規定による <input type="checkbox"/> 書面 <input type="checkbox"/> 口頭 <input type="checkbox"/> 撮影代表者判断
24	目録公開と利用条件の 確認方法 (備考)	<input type="text"/>
【調査撮影申請書・許可証・利用条件確認書などのアップロード】 <small>※申請書・許可証・確認書などはまとめて個別にアップロードしてください。(ファイル形式(Word(.docx)、Excel(.xlsx)、PDF(.pdf)などは任意です。) またファイル名の冒頭に"20220401"のように発給または受領の年月日を記入してください。</small>		
25	関連ファイル1	<input type="text" value="※ファイルを選択(ここにドロップすることもできます)"/> <input type="button" value="参照"/> <input type="button" value="取消"/>
	関連ファイル1 (備考)	<input type="text"/>

International Image Interoperability Framework (IIIF)

■ IIIF : とりふるあいえふ

- 国際的なデジタルアーカイブの相互運用の枠組み

■ 画像が主な対象

- 動画や3DCGなども実は対象

■ 利点

- デジタルアーカイブとのやり取りを共通化 : デジタルアーカイブ同士の相互運用をスムーズにする
- 画像の閲覧, 比較, 操作, 注釈など操作にも共通化していく

■ IIIFの機能

- 配信 : IIIF Image API
- 表現 : IIIF Presentation API
- (認証・セキュリティ : IIIF Authentication API)
- (検索 : IIIF Search API)
- (追加・更新・削除 : IIIF Change Discovery API)
- (状態 : IIIF Content State API)

- 史料編纂所の画像配信サーバは Image API と Presentation API の2つを実装している。

■ 史料編纂所での対応

- 2015年12月より開始 : 国内では最古参(たぶん)

史料編纂所所蔵史料のデータ利用条件

1. 史料編纂所が所蔵する原本等の史料画像データ

- 史料種別(区分)が「貴重書(原本・古写本類)」・「特殊蒐書」(一部を除く)・「写本」であればCC BY相当
- “オープンデータ”

2. 史料編纂所出版物の版面画像データ

- 史料種別(区分)が「本所出版物」の画像データはCC BY-NC-SA相当

3. 上記以外の史料等画像データ

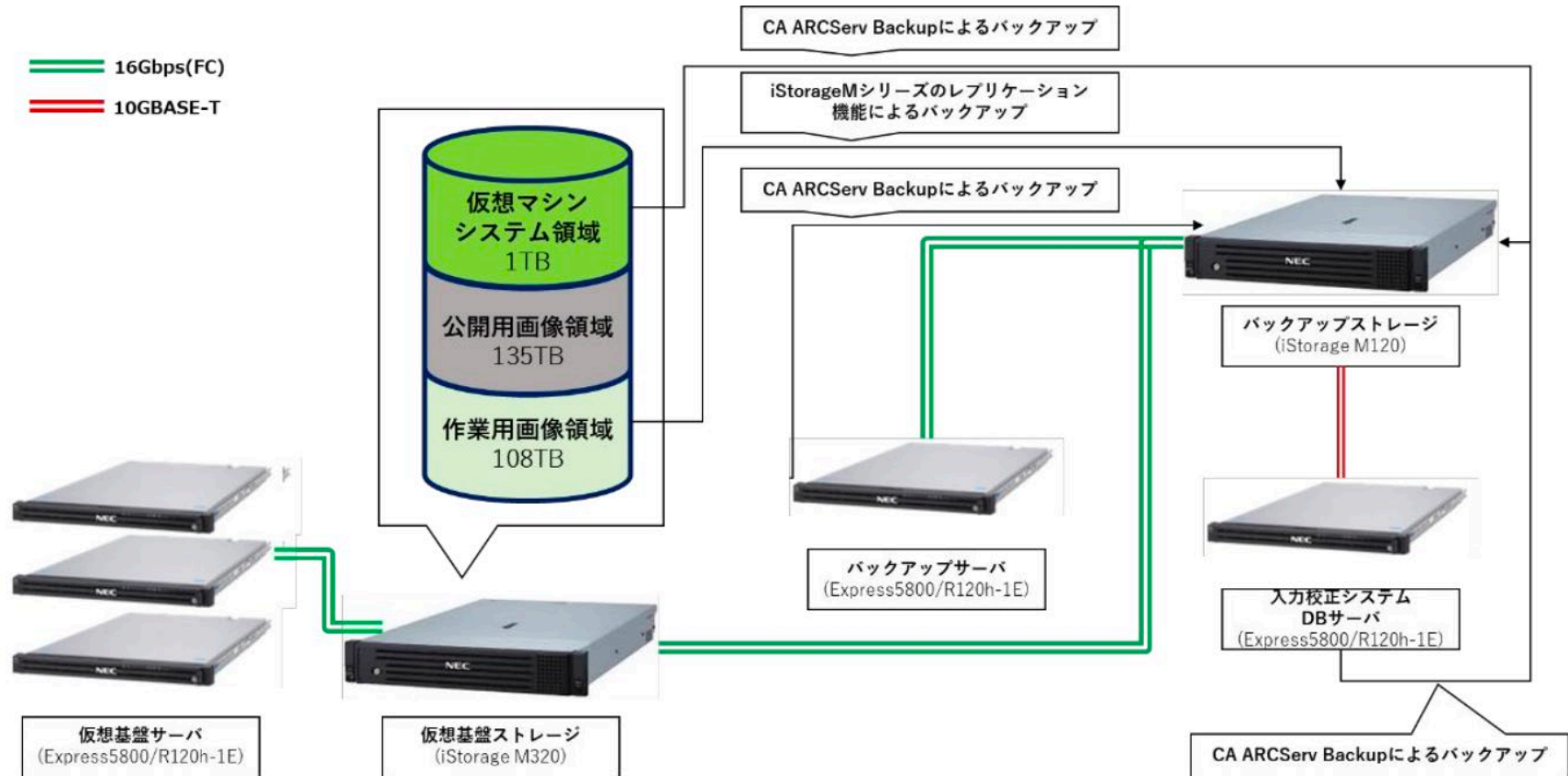
- 史料編纂所図書室へ申請

■ 上記は2019年4月1日より開始

■ <https://www.hi.u-tokyo.ac.jp/faq/reuse>

- 史料編纂所所蔵以外の史料
 - 史料所蔵者が設定. それを表示

データのバックアップ



バックアップデータ:

- ・システム・データベース・ウェブサイト・NAS(共有ファイル)等:2世代
- ・史料画像:1世代
 - 作業用画像(RAW->jpeg)をバックアップ, 公開用画像は対象外
 - (実際はRAWも保存 → “写真室の努力”)

遠隔データバックアップ

■きっかけ: 2011年東日本大震災

■どこへ?: 京都大学東南アジア地域研究研究所(地域研究情報統合センター)

- 原正一郎教授のご尽力による
- 2012年から開始. 現在も継続.

■方法

- バックアップデータ作成 → HDDへデータバックアップ → これを送る
 - 当初は人が運ぶ. 今は宅急便.
- 頻度: 年1回

■作成

- 差分のバックアップデータ作成
- 基本はデータベースと史料画像
- システムリプレースのタイミングでフルバックアップ