欧州原子核研究機構（CERN）における研究データ整備および欧州学術機関におけるオープンアクセスの取組み

平成29年度国立情報学研究所実務研修報告会

2017年12月27日

東京大学附属図書館 情報管理課
横井慶子

目次

- CERNとは
- INSPIREでのデータ整備
- TIB訪問調査
- MPDL訪問調査

CERN

英語：European Organization for Nuclear Research

仏語：Conseil Européen pour la Recherche Nucléaire

欧州原子核研究機構（CERN）

- 1954年発足。世界最大規模の素粒子物理学の研究機関
- 加盟国：ヨーロッパ22カ国　※日本はオブザーバー国
- 年間予算：約1,200億円
- 職員：約2,500名
- ユーザー：12,816名（世界77カ国）　※2017/1/12時点

国別CERNユーザー数　(2017/1/12時点)

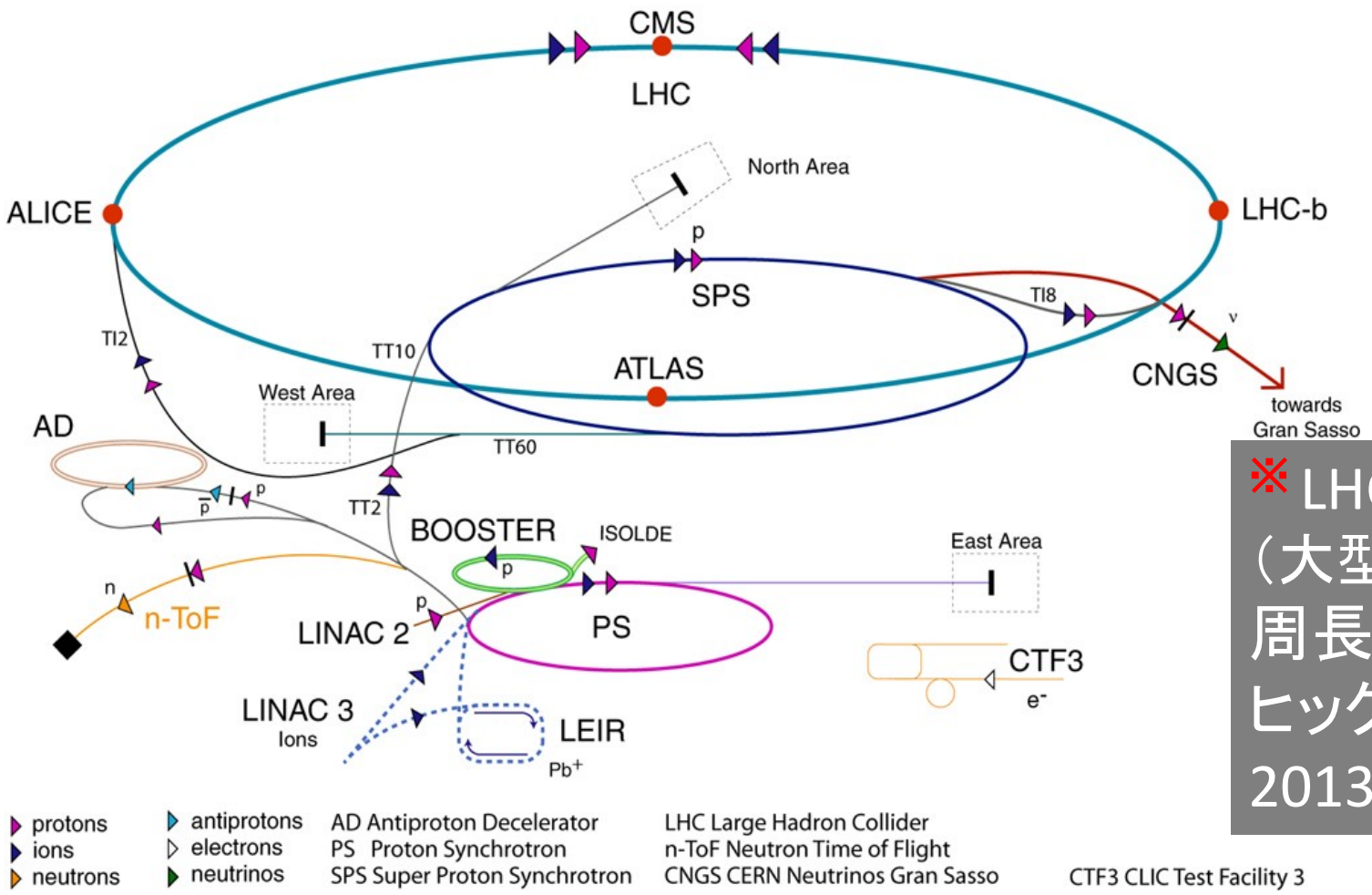
- 加盟国: 60 %
- オブザーバー国:26 %

（うち日本 2.3 %）

国名	ユーザー数
アメリカ	2,018名
イタリア（加盟国）	1,564名
ドイツ（加盟国）	1,343名
ロシア	1,046名
フランス（加盟国）	950名
イギリス（加盟国）	942名
中略	
日本	294名

CERN内の加速器

地下100m付近に複数の加速器があり
LHC※への入射ビームをつくる。



※LHC：Large Hadron Collider
（大型ハドロン衝突型加速器）
周長26.6km のCERN最大の加速器。
ヒッグス粒子の存在を確認に成功。
2013年ノーベル賞受賞

図は「CERNの概要」より引用
http://atlas.kek.jp/public/IntroductionOfCERN.pptx

CERNでの実験への日本の参加

LHC

- **衝突実験**

✓LHC加速器建設（KEK）

✓ATLAS実験(KEK,東大など16機関)

✓ALICE実験(広大、東大、筑波大）

✓LHCf実験（名大、早大ほか）

✓LEP加速器によるOPAL実験

(2000年に終了 東大,神戸大)

SPS、PS

- **固定標的実験**

✓COMPASS実験(山形大他)

✓ASACUSA実験(東大他)

SPS→伊グランサッソ

- **ニュートリノビーム**

✓OPERA実験(名古屋大他)

INSPIREでのデータ整備

INSPIREとは

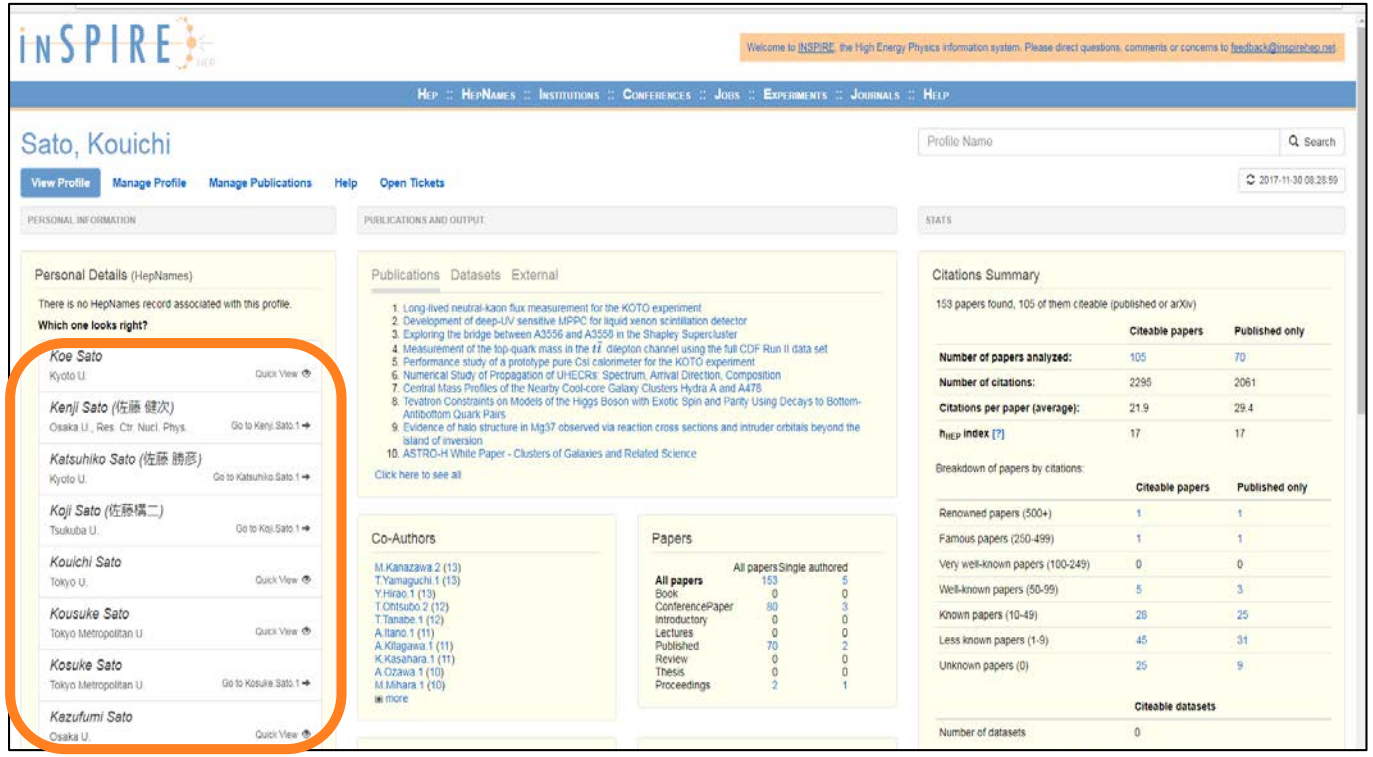
- 高エネルギー物理学分野の文献データベース
- 高エネルギー物理学分野の研究機関CERN, DESY（Deutsches Elektronen-Synchrotron:ドイツ）, Fermilab（アメリカ）, IHEP（Institute of High Energy Physics Chinese Academy of Science：中国）, and SLAC（SLAC NATIONAL ACCELERATOR LABORATORY ：アメリカ）による国際共同運営

INSPIREを構造するレコード

- HEP（論文）
- HEPNames（著者）
- Institutions（機関）
- Conferences（会議）
- Experiments（実験）
- Journals（雑誌）

自動分析、生成 → **Author Profile**
(著者ごとの論文プロファイル)

著者候補



研修内容（On the Job Training）

- 業務１：著者情報と業績情報の結びつけ
- 業務2：著者情報(HepNames)の充実化

対象

「tokyo」を名称に含む機関（東大、東京理科大等）に

<u>在籍歴のある</u>著者約650名

業務1.著者情報と業績情報の結びつけ

Author Profile > Management publications

❖異なる著者の業績が混在

❖業績ごとに以下の情報を掲載＝判定の手がかり

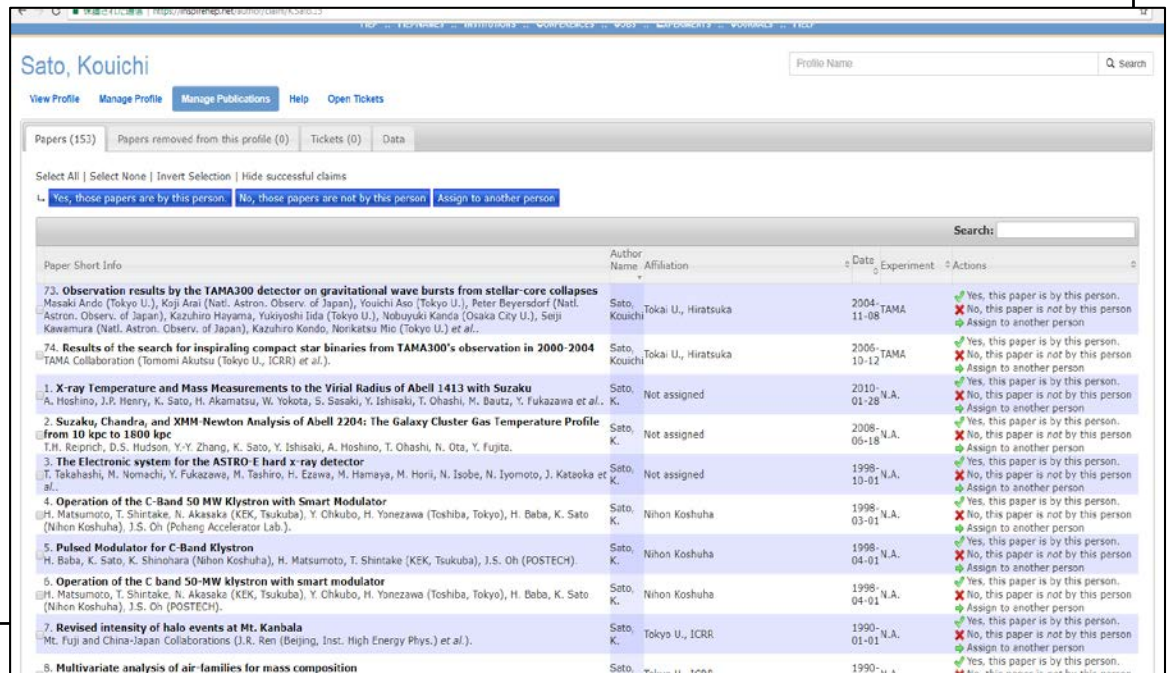
・Paper Short Info(論題、共著者名)

・Author Name（著者名）

・Affiliation （著者所属）

・Date（業績公表年）

・Experiment（プロジェクト名）



所属機関情報＋業績公表年=業績特定の最大の手がかり

Paper Short Info	Author Name	Affiliation	Date	Experiment
73. **Observation results by the TAMA300 detector on gravitational wave bursts from stellar-core collapses** Masaki Ando (Tokyo U.), Koji Arai (Natl. Astron. Observ. of Japan), Youichi Aso (Tokyo U.), Peter Beyersdorf (Natl. Astron. Observ. of Japan), Kazuhiro Hayama, Yukiyoshi Iida (Tokyo U.), Nobuyuki Kanda (Osaka City U.), Seiji Kawamura (Natl. Astron. Observ. of Japan), Kazuhiro Kondo, Norikatsu Mio (Tokyo U.) *et al.*.	Sato, Kouichi	Tokai U., Hiratsuka	2004-11-08	TAMA
74. **Results of the search for inspiraling compact star binaries from TAMA300's observation in 2000-2004** TAMA Collaboration (Tomomi Akutsu (Tokyo U., ICRR) *et al.*).	Sato, Kouichi	Tokai U., Hiratsuka	2006-10-12	TAMA
1. **X-ray Temperature and Mass Measurements to the Virial Radius of Abell 1413 with Suzaku** A. Hoshino, J.P. Henry, K. Sato, H. Akamatsu, W. Yokota, S. Sasaki, Y. Ishisaki, T. Ohashi, M. Bautz, Y. Fukazawa *et al.*.	Sato, K.	Not assigned	2010-01-28	N.A.
2. **Suzaku, Chandra, and XMM-Newton Analysis of Abell 2204: The Galaxy Cluster Gas Temperature Profile from 10 kpc to 1800 kpc** T.H. Reiprich, D.S. Hudson, Y.-Y. Zhang, K. Sato, Y. Ishisaki, A. Hoshino, T. Ohashi, N. Ota, Y. Fujita.	Sato, K.	Not assigned	2008-06-18	N.A.
3. **The Electronic system for the ASTRO-E hard x-ray detector** T. Takahashi, M. Nomachi, Y. Fukazawa, M. Tashiro, H. Ezawa, M. Hamaya, M. Horii, N. Isobe, N. Iyomoto, J. Kataoka *et al.*.	Sato, K.	Not assigned	1998-10-01	N.A.
4. **Operation of the C-Band 50 MW Klystron with Smart Modulator** H. Matsumoto, T. Shintake, N. Akasaka (KEK, Tsukuba), Y. Ohkubo, H. Yonezawa (Toshiba, Tokyo), H. Baba, K. Sato (Nihon Koshuha), J.S. Oh (Pohang Accelerator Lab.).	Sato, K.	Nihon Koshuha	1998-03-01	N.A.

整備前のデータ(イメージ):Koichi.Satoの場合

手がかり:著者名・掲載年・所属のみ！

著者データ

Koichi. Sato
(Tokyo U.)

Koji. Sato
(Osaka U.)

K. Sato
(Tokyo
Metropolitan U.)

氏名も所属も一致
2015年に東大にいた？

京大に在籍歴有？
OR
同姓同名の別人？

著者の可能性あり

論文データ群
(異なる複数著者の論文情報も含む)

論題	掲載年	著者名	著者所属
A	2015	Koichi.Sato	Tokyo U.
B	1988	Koichi.Sato	Kyoto U.
C	1999	K. Sato	Tokyo U.
D	2017	K. Sato	Tokyo U.
E	2009	K. Sato	NA(不明)

候補著者全てに著者である可能性あり

東大に在籍歴があればC,Dの著者の可能性あり

著者データの情報（HepNames由来）

- **アルファベット表記の氏名**

※ファーストネームはイニシャルのみの場合も多い

- **所属機関名**

※いつの時点の所属情報か不明。院生時点？教員時点？

⇒そもそも対象著者の同定（漢字表記）が必要

⇒そのうえで所属歴の特定が必要

著者同定＝漢字フルネーム
各Webソースの特性に応じて検索

- CiNii Articles　:収録対象広い。院生も対象。
- Researchmap　:
- KAKEN　:

一定の研究歴がある著者を特定しやすい。
ノイズが少ない。

- NDL-OPACの博士論文検索　:収録年数が広い
- CiNii Dissertation　:様々な検索条件を指定可能、ノイズ少ない。
- JAIRO/各機関リポジトリ:最新の学位取得者情報も含む
- Google　:様々な検索式で検索（機関名、指導教員名など活用）

著者情報の公開状況の傾向

- Researchmap
- KAKEN

現役研究者に有効
(科研費取得経験有り、研究歴あり)

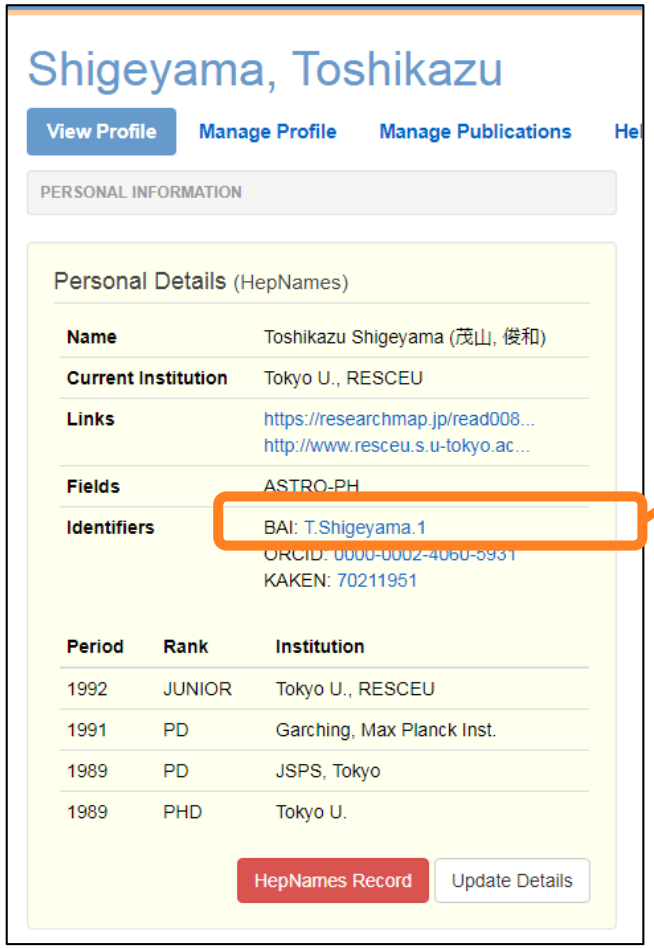
- 著者個人のウェブサイト
- SNS（Facebook, Researchgate, Linkedin）

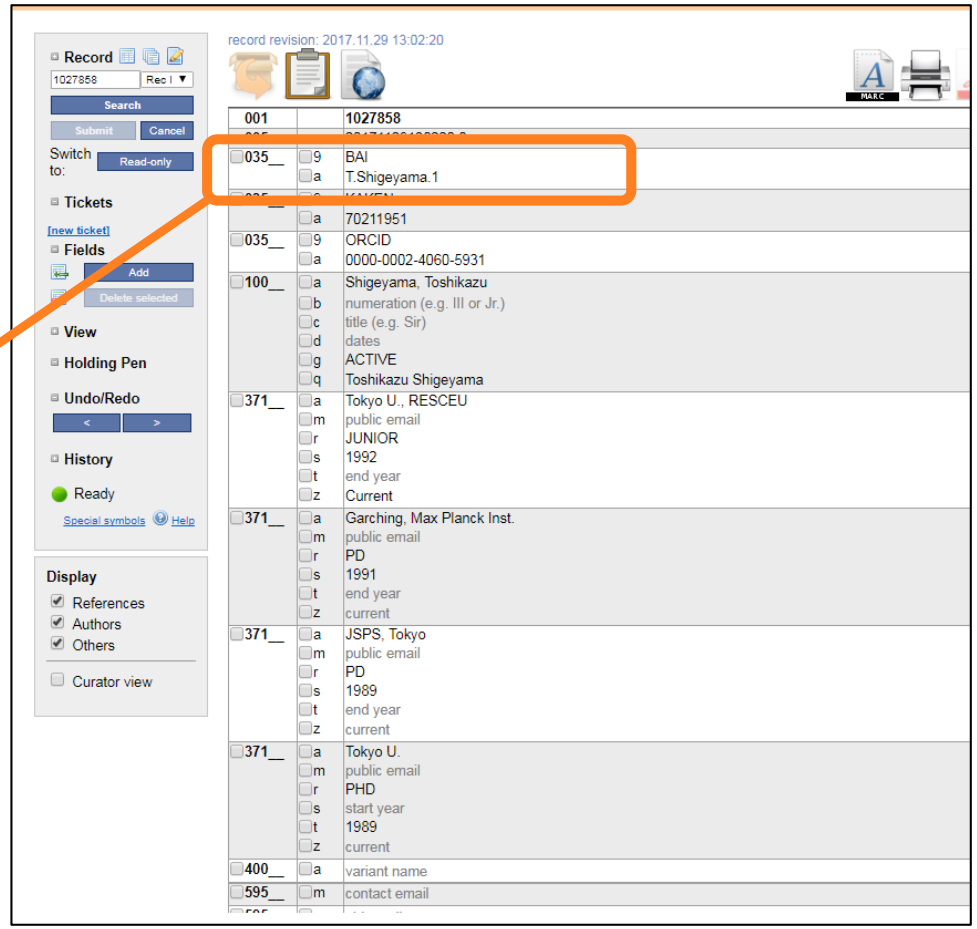
若手、特に院生に有効
Linkedinは研究活動をやめて別の業種に移った人に多い

著者情報の特定困難
・研究活動をやめて相当数の年数が経つ研究者
・日本で教育を受けたが、海外で職を得た研究者

業務2:著者情報の充実化

Author profile⇔BAI ⇔ HEPNames





著者情報を追加

- ORCID ID
- KAKEN ID
- 漢字氏名
- 異動歴
- ウェブサイト
- 学位の有無
- 指導教員名などなど

001		1027858
005		20171129130220.0
035__	9	BAI
	a	T.Shigeyama.1
035__	9	KAKEN
	a	70211951
035__	9	ORCID
	a	0000-0002-4060-5931
100__	a	Shigeyama, Toshikazu
	b	numeration (e.g. III or Jr.)
	c	title (e.g. Sir)
	d	dates
	g	ACTIVE
	q	Toshikazu Shigeyama
371__	a	Tokyo U., RESCEU
	m	public email
	r	JUNIOR
	s	1992
	t	end year
	z	Current
371__	a	Garching, Max Planck Inst.
	m	public email
	r	PD
	s	1991
	t	end year
	z	current
371__	a	JSPS, Tokyo
	m	public email
	r	PD
	s	1989
	t	end year
	z	current
371__	a	Tokyo U.
	m	public email
	r	PHD
	s	start year
	t	1989
	z	current
400__	a	variant name
595__	m	contact email

その他1：古いデータの削除

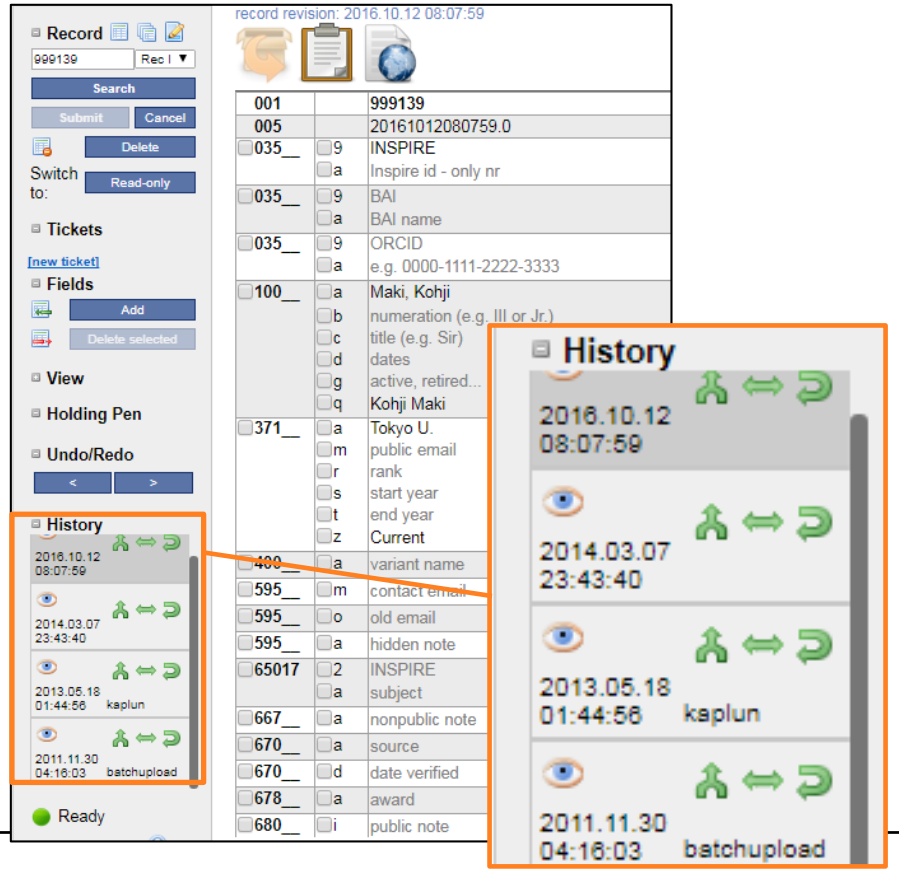
HEPNamesはあるが論文データがない場合…

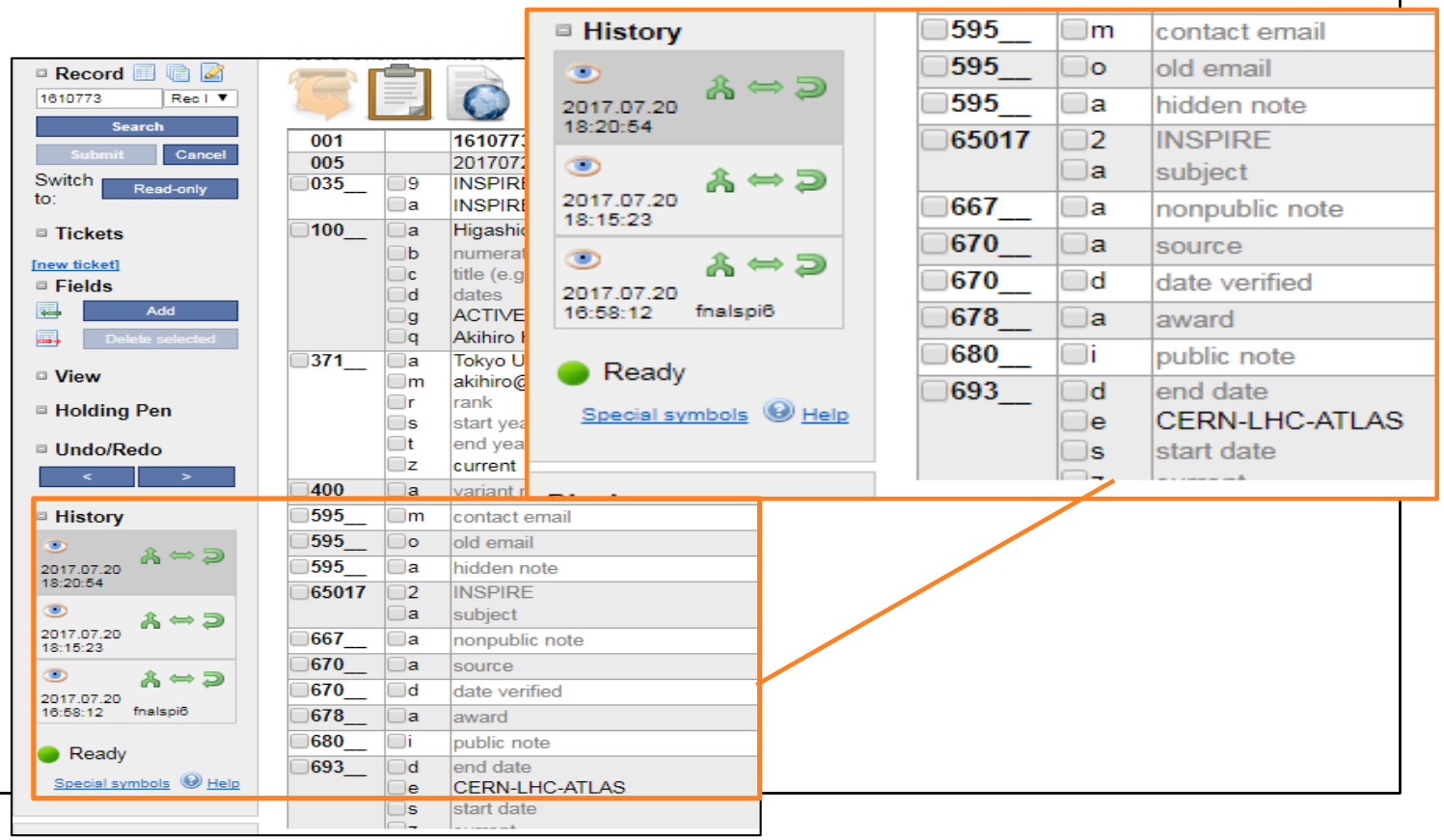
＜パターン1＞

2011年に一括生成 → **削除**

＜パターン2＞

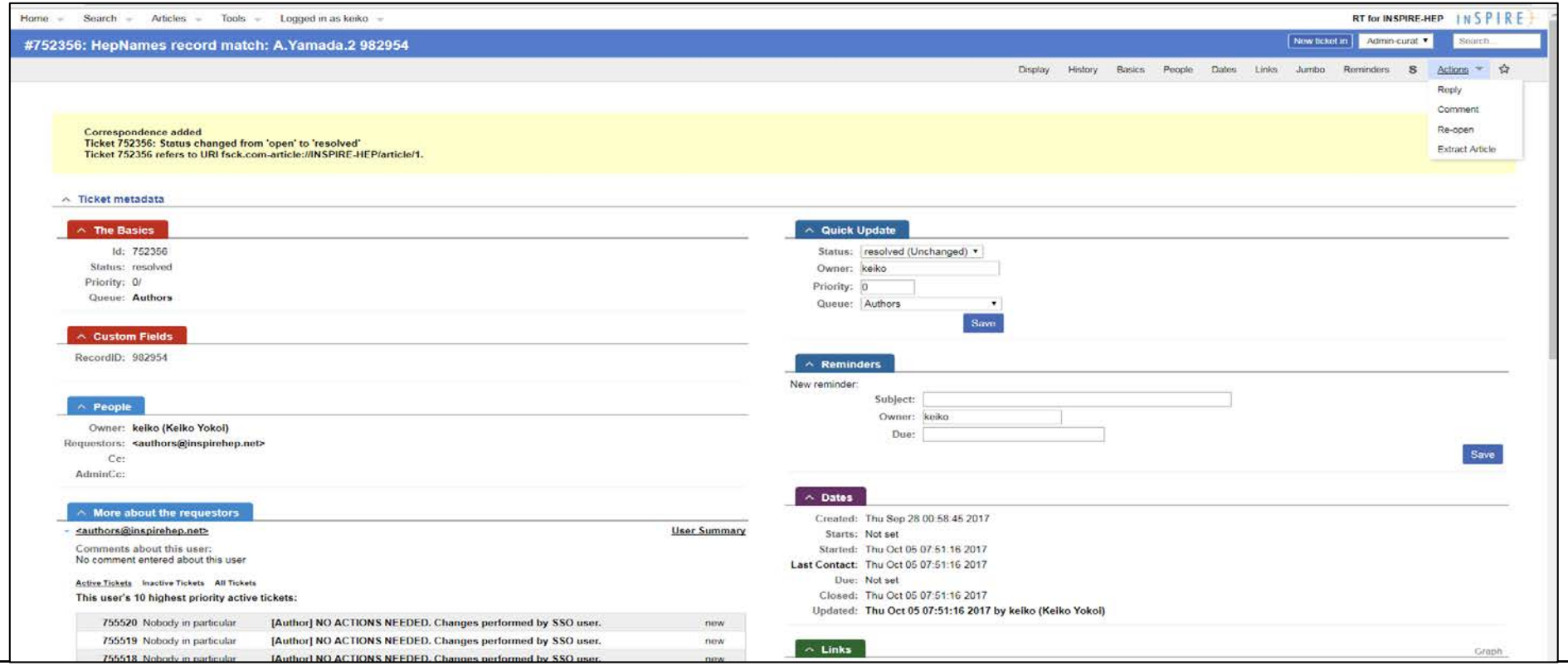
プロジェクト単位で生成
論文はこれから（見込） → **保留**





その他2：
著者からの申請を受けてデータ整備

チケット管理画面



その他：
日本でのORCID普及のため資料づくり

- INSPIREとORCIDを連携、同期の手順を和訳

情報源：https://inspirehep.net/info/HepNames/connect_orcid

- INSPIREからORCIDへの業績データのインポート手順を和訳

情報源：

https://inspirehep.net/info/faq/general?ln=en#CV_Citesummary_BibTeX_output

https://support.orcid.org/knowledgebase/articles/1807549-importing-works-from-a-bibtex-file

今後の課題

- 正確なデータをいかに維持するか（新しい業績データ日々追加）
- ORCID の活用

リプレイス後のINSPIREでは、個人向けサービス利用時にORCID IDをログインに使用

← ORCID の普及推進

- 研究者（学会？）とのコミュニケーション
- キュレーターに物理学の知識

TIB訪問調査

ドイツ国立科学技術図書館

（TIB: Technische Informationsbibliothek）

RADAR: Research Data Repository

- ロングテールに属する多種多様な分野向けのデータリポジトリ
- DFG助成(2013-16)を経て，2017年3月始動
- ４段階プロセス：Pending(編集可，6ヶ月)，Review(編集不可，6ヶ月), Archive（RADAR ID付与，5-15年）, Publish（DOI付与，25年以下）
- ビジネスモデル

  Archival: 500 €（年間使用料）+ 1GBごとに毎年0,39 € (正価)

  Publication: 1GBごとに6,37 €　25年は保証 (正価)

MPDL訪問調査

Max Planck Digital Library

論文生産数調査の概要

- ScopusとWoSからの購入データをData Analysis team（5名）が分析。PostgreSQLとPYTHONのスキル要。

- 調査対象：Max Planck所属研究者のほか，

スイスや韓国の研究機関，マサチューセッツ工科大，

カリフォルニア工科大なども。

→SCOAP^{3}の影響で，2014年以降OA論文数が増加

→Springer Compact契約(2016-)でSpringerのOA論文数も増加

オフセット契約上のリスク，注意点

- OA論文数が契約範囲数を超えると支払額増のリスク

←過去10年に対象出版社の雑誌に掲載された自機関研究者の論文数の推移を見て，適切な予測を立てる！

- 出版社との交渉では，その出版社の雑誌掲載論文全体のうち，自機関研究者の論文数が占める割合を把握