

名寄せとID

森 雅生・東京工業大学

名寄せ

データ分析でよく顕現する課題。

例

- 論文書誌情報の著者名（同姓同名）
 - ✓ 共著などを排除し研究業績数（法人評価）
- 企業名（正式名称の表記揺らぎ）
 - ✓ 卒業生の就職先（就職支援）
- 大学・機関名（特に海外、和名英名揺らぎ）
 - ✓ 国際連携事業、派遣留学生滞在先

ID

- 名寄せを避ける ← IDの導入。
- 国内のIDはなんとかなるが、国外のIDは？
 - ✓大学名や企業名
- 個人のIDは、個人情報保護が問題に。
 - ✓共済職員番号、学籍番号
- IRで活用できる可能性のあるIDについて紹介。
 - ✓PID, Persistent Identifier

個人のID

- 個人情報保護の問題
 - ✓公開を前提としたIDを使いたい
- ResearcherID
 - ✓Clarivate Analytics社によるproprietary ID
 - ✓研究者を対象
- ORCID (Open Researcher and Contributor ID)
 - ✓独立系非営利団体によるPersistent ID
 - ✓研究者、その支援者を対象（学生もOK）
 - ✓論文情報、獲得資金情報、職歴・学歴、所属などの情報も集約できる。

国・都道府県・地名など

- 国・言語

- ✓ ISO 3166 …… 2文字、3文字、番号表記あり。
- ✓ Pythonなどライブラリで取得可能、ボランティアがメンテナンス。
- ✓ JIS X304に日本語表記あり（ただし有償）。

- 都道府県・地名

- ✓ 都道府県番号（JIS X401）
- ✓ 郵便番号 <https://www.post.japanpost.jp/zipcode/dl/roman.html>
 - ✓ なぜかLHAで圧縮されている。
- ✓ 地名 <https://www.kodokensaku.mlit.go.jp/motas/>
 - ✓ 国土交通省 自動車登録関係コードシステム

組織等のID

- 国内大学等

- ✓ 文部科学省 https://www.mext.go.jp/content/1422972_03.pdf
- ✓ JSPS（科研費） <https://www-kaken.jsps.go.jp/kaken1/kanList.do>
- ※ 政府官庁が事務処理目的で管理しているので、その意味で一応PID
- ※ 他のIDとの連携が薄い

- 海外の大学研究機関

- ✓ GRID（Global Research Identifier Database）大学、研究機関、企業、助成財団など。運営は商業DB系。
- ✓ ROR（Research Organization Repository）同上、運営はORCID系。
- ※ 両者は連携して運営されることが決定。
- ※ apiや他のIDとの連携が厚い
- ※ Wikidataを利用して日本語名が取得可能

組織等のID

- 高等学校

- ✓ 大学入試センター <https://www.dnc.ac.jp/albums/abm00007009.pdf>
- ✓ おやおや、9月入学の話と連動してますかい？

学校コードの取り扱いに関する意見募集の実施について

令和2年4月28日

総合教育政策局調査企画課

文部科学省では、従来より学校基本調査等の実施にあたり学校（専修学校及び各種学校を含む。以下同じ。）を識別するため学校調査番号を用いてきたところですが、これについて都道府県ごとに独立して番号が振られていることや、番号の不足のために廃校となった学校の番号を新設校に流用するなどの状況があることから、全国的あるいは経年で見ると、同じ学校調査番号が割り振られた学校が複数存在する状況が見られました。

このことを踏まえ、新たに全国の学校に一意の「学校コード」を付与するとともに、学校に一旦付与された学校コードは変更せず、他の学校には流用しないことを基本とすることにより、上記課題の解決を図ることとしました。

さらに「学校コード」については、広く社会に公開（全ての学校の学校コード及びその属性情報を文部科学省のホームページで公表予定）することとすることにより、学校基本調査のみならず各種の調査研究などにおいて活用されることを想定しています。これにより、政府の第3期教育振興基本計画に定める、客観的な根拠を重視した教育政策の推進に

組織等のID

- 企業

- ✓ 法人番号（国税庁）

- <https://www.houjin-bangou.nta.go.jp/website/riyo-hoho/>

- APIも公開。他のIDとの連携は薄い

- ✓ GRID、ROR

- 大手企業が掲載されている。日本語名は無し。

その他のソース

- Wikidata

- ✓ Wikipedia, Wikimedia commonsで利用されている構造化データ。
- ✓ APIで利用可能。
- ✓ 各種IDが掲載。
- ✓ 多言語での表記。
- ✓ 国際的なPIDで日本語表記をここに求めることができる。
- ✓ CC0 (Creative Commons 0) の元、利用可能。