

## CA1975 XXXXXXXXXX オルタナティブな情報を保存する:統計不正 問題からこれからの図書館を考える

ふくしまゆきひろ  
福島幸宏\*

### はじめに

2018年12月に発覚した厚生労働省所管の毎月勤労統計調査の不正問題は<sup>(1)</sup>、問題が経済・雇用政策の根幹にかかわるだけに、大きな波紋があった。非常に多くの検証や報道がなされたなか<sup>(2)</sup>、明治初年以來の伝統を誇る日本統計協会では、月刊誌『統計』において、2019年中に5回も「統計の信頼性向上をめざして」という特別企画を組んでいる。結局この問題を発端にした政府統計の不備は、「政府の一斉点検によると、56ある基幹統計のうち約4割で問題が見つかった」<sup>(3)</sup>とされている。日本の基幹統計が全く信頼性を失ったことのインパクトは非常に大きい。

この事件は、一見直接的な当事者でない図書館にも深い関係を持つ。従来、図書館は収蔵している個別の図書館資料の記述が検証を要することを前提に資料提供を行ってきた。しかし、2000年代前半以降、図書館機能の再構成が検討されるなかで重視されてきたビジネス支援などの課題解決支援サービスや<sup>(4)</sup>、その後に広く議論され実践されてきているオープンデータ、オープンガバメントを巡る動向のなかでは<sup>(5)</sup>、前述の資料提供の前提は意識しつつも、踏み込んで資料を提示・データ化していく以上、より信頼できる資料を中心に行うことは無意識の前提となっていたのである。

今回、特に信頼度の高い情報源とされていた政府統計について、上記の状況が明らかになった。レファレンスをも含めたこの間の図書館の取組は、砂上に楼閣を営々と築いていたと言われても反論できるだろうか。

そして、これまで「確か」と思われていた情報の信頼性が揺らいだ時代において、図書館の役割をどのように考えればよいのだろうか<sup>(6)</sup>。

### 近代日本と不全な統計

もっとも、近代の各種統計の数値はより慎重に扱われるべき、という議論が統計学の分野からも提起されている。以下、佐藤正広の一連の研究から指摘する<sup>(7)</sup>。まず、統計院書記が、1884(明治17)年に西日本の各地の、特に農作物統計を調査した記事から、当時の統計調査の実態として、地域の事情に詳しい人に頼った概算、平年作の状況の聴取に基づく推計、前年と今年との比較からの推計、などの手法があることを指摘している。

これらは、佐藤も指摘するように、中央政府もまだまだ脆弱で、地域においては近世以來の構造を受け継がざるを得ない明治前半期の、特に農業統計に関わる部分においては、仕方がない部分もあるかもしれない。

しかし、近代化の要のはずの海運の状況を直接把握するために作成される港湾統計が、1920(大正9)年の段階になっても正確にカウントしたものではない実態が、当時の国勢院が各府県に対して行った調査から浮上する。最終的に『大日本帝国港湾統計』にまとめられる各種の基礎数値が、不確実な根拠に基づいているというのである。例えば北海道では、「最近5年間出入船舶輸出入貨物噸数価額」の調査に際し、突然かつ5年間にさかのぼる調査のため、小樽港などを参考に全道の数値を机上で推計したと報告している。また、農商務統計の生産統計のほとんどが「揣摩臆測」「筆なめ」「達観調査」であるとされている。この要因には、通常扱う事務以外の範囲での数値提出の要請、調査項目の定義の不正確さ、調査方法の不明確さ、中央官庁による安易な照会の多発によるモラルの低下、などが指摘されている<sup>(8)</sup>。

しかし、佐藤は、推計であっても「同時代で同じ道府県にいる人から見て、不自然と思われぬ程度の確からしさ」を持つものであり、「非標本誤差」に十分注意しながらもこれらの統計を使うべきと主張する。そして、この課題の解消には訓練された人員と仕組みとが必要で、それらは1920(大正9)年の初の国勢調査から段階的に整備され、戦後改革によってある程度達成されたとの見通しを述べている。言い換えれば、国家が国民生活を直接把握することが必須となった総力戦段階に立ち至って、より正確な統計が整備されようとした、と言えるのであろう。

しかし、意図的な操作のモメントとその結果は当然払拭されない。アジア太平洋戦争の敗戦時に、占領軍に提出する戦略物資に関する統計を日本側で操作しようとしていたことがわかっている。敗戦直後の1945(昭和20)年8月25日、近畿地方総監府から管下の各府県知事宛に「諸統計調査ニ関スル件」が通達された。これは、占領軍からの行政資料提出の要求に備えて、関係各機関の「諸統計ヲ統一整備」するために出された指示であったが、実は異なる形式のものが2通発出されている。より詳しいものには、○統計の表に「概数」と「誤差修正」の2つの欄を設け、「概数」欄には真実数をあげ、「誤差修正」欄には占領軍に提示すべき数字を記入すること、○関係機関から占領軍に提示する数字は、「誤差修正」欄の数字になるので、提示用統計表を至急作成し、その他の統計は焼却および隠匿すること、○この通知は焼却すること、などが指示されていた。実際にこの指示に基づいて京都府から報告された数字を検査すると、医師数は、「真実数」は京都府内全体で2,141人であるの

\*東京大学大学院情報学環

に対して「誤差修正」欄は1,065人となっている。他の数値もほぼ同様の傾向があり、ほとんどの項目でほぼ半減になっているのである。もっとも1947(昭和22)年に厚生省公衆保健局が発行した『衛生年報 昭和16～20年』には、昭和20年度末の京都府の医師数として2,663人が計上されている。その間ほどのように公式な数字を操作していたか判然としないが、内部に資料が遺されていたことによって、最終的にはほぼ正確に把握されたものと考えられる<sup>(9)</sup>。

結局、近代の統計は不全なものとして受け止め、利用していく必要があるのである。

### これからの図書館や関連機関のあり方

もっとも、この統計の不全は、図書館界においてもその背景の構造も含め、前述と同様の問題を今も惹起している。日本図書館協会が毎年度取りまとめている『日本の図書館 統計と名簿』は特に公立図書館を議論するうえで基本となる資料だが、数字のブレがあることはよく知られている。文部科学省の委託調査の報告書でも「自治体によって登録者の範囲、登録の有効期限など統計の取り方がまちまちなので、自治体ごとの比較は困難である」<sup>(10)</sup>と表現されたりしている<sup>(11)</sup>。明治の吏員たちが統計調査の負担を軽減するためにとった様々な便法は、統計を含めた各種の調査に対応している現在の図書館現場に共通するのではないだろうか。そして、この統計の不揃いの問題は、解決策が見出しにくいために、課題と認識されながら、比較的等閑視されてきたのではないか。しかし、新たなサービスへの注目や運営効率化の議論のなかで、努力義務として図書館評価が2008年に改正された図書館法第7条に示されたように、こと図書館においても、統計をどのように作成していくかが、重要な課題となってきている<sup>(12)</sup>。

これらの課題は別途解消されてくるとして、この図書館自体の状況も含め、統計がそのようなものであると観念するとした際、当面どのように対峙していけばよいのだろうか。

「デジタルアーカイブ」という言葉を創ったとされる月尾嘉男は、おそらく今後永遠に続く爆発的な情報量の増加という事態のなかで、トランプ政権以降の米国における“alternative fact”や“deepfake”などを例に挙げながら、「急速に技術が進歩する情報社会でデジタルアーカイブがどのような問題に直面しているかを知ることが重要」と指摘する<sup>(13)</sup>。これに対し、この月尾の講演をレポートした鷹野凌は、私見と断りながら、「仮にフェイク情報だとわかっている、可能な限りアーカイブしたほうがいいと思う」「あらかじめ「これはフェイクだ」というメタデータを付与しておけばいい。あとからフェイクと判明した場合も同様、消去するので

はなく、「これは間違っている」「理由はこの情報を参照すればわかる」といったメタデータを付与する」「というのは、そのとき「間違っている」と判断したことが、間違っている場合もあるのだ」「そういう意味では、フェイクのアーカイブにも価値はあるのだ」と述べる<sup>(14)</sup>。この対話は非常に示唆的である。

前述の敗戦時の統計操作の挿話と京都府の「真実数」をなぜ我々は認識し得ているのか。これは、京都府行政文書「近畿地方総監府諸統計調査ニ関スル件併ニ人口調査表各種」(昭20-68)というファイルが京都府に保管され、そこに本来は「焼却」と指示されていた通知とそれに対応した処理過程が綴られていたためである。つまり、その時点の日本政府としては公式には「間違っている」と言わざるを得ない処理と数字が、資料が遺されていたために跡付けることが可能となったのである。

結局、その時点時点の「正確さ」には留意しつつも、後生も含めた多様な眼の検証に耐えるために、オルタナティブの数字を遺す、処理の過程を遺す、データを捨てずに保存しておく、という機能を社会のどこかが担わなければならないのであろう。

国レベルではまだ議論の余地があるが、多くの地方公共団体や大学では、集められる限りのデータを集め、捨てずに保存しておく機能は図書館が担わざるを得ない。冊子体にまとめたものを紙や電子で収集するという従来の発想を改め、収集範囲を広げて関係資料をごっそりと収集する、その蓄積のためにデジタル技術を正面から導入するということがまずは考えられるだろう<sup>(15)</sup>。そして、一方では、地下茎のようにひそかに関係各所や個人に渡りをつけて、資料と情報を蓄積することが重要かもしれない。国立国会図書館の憲政資料室や米国の大統領図書館を持ち出すまでもなく、意思決定の過程は必ずしも公的文書に残るとは限らないからである<sup>(16)</sup>。

### おわりに

2020年の年初から始まった新型コロナウイルス感染症の感染拡大状況下においても、公式に発表される数字がどのように作成されているか、その背景を知悉するリテラシーを市民は常に求められている<sup>(17)</sup>。この状況は、100年前のスペイン・インフルエンザの際と、実は大きくは変わらない<sup>(18)</sup>。

図書館の特性が、その資料と機能のロングテールにあるとすれば、これらの状況が事後に十分検証できるように、各種の情報を積極的に収集するよう、制度と心性が更新されなければならない。また、事後の検証を社会のどこかで担保することによって、今まさに各種の課題に直面している現場の負担を減らしていくことにもつながると考える<sup>(19)</sup>。そして、冒頭に述べた図書館

固有の文脈に戻しても、より深い課題解決支援サービスやオープンデータの提供につながる事が期待される。その際の営為は、つとに主張されている「図書館の態度を問う‘図書館の社会的責任’の理念は、価値中立ではありえず、進歩的価値観(progressive priorities)の採用を迫る」<sup>(20)</sup>ということを念頭に行われるべきであろう。

(1) 厚生労働省. 毎月勤労統計調査において全数調査するとしていたところを一部抽出調査で行っていたことについて. [https://www.mhlw.go.jp/stf/newpage\\_03207.html](https://www.mhlw.go.jp/stf/newpage_03207.html), (参照 2020-05-24).

(2) 朝日新聞社の「論座」では「統計不正問題が意味するもの」として、10本の特集記事を掲載している。他の報道機関でも同様に重視された。<https://webronza.asahi.com/feature/articles/2019021300008.html>, (参照 2020-05-24).

(3) 前掲。

(4) 全国公共図書館協議会編. 2015年度(平成27年度) 公立図書館における課題解決支援サービスに関する報告書. 全国公共図書館協議会, 2016, 68p. <https://www.library.metro.tokyo.jp/pdf/zenkouto/pdf/2015all.pdf>, (参照 2020-05-24).

(5) 例えば、『情報の科学と技術』では、2015年12月号で「特集:オープンデータ」を組んでいる。[https://www.jstage.jst.go.jp/browse/jkg/65/12/\\_contents/char/ja](https://www.jstage.jst.go.jp/browse/jkg/65/12/_contents/char/ja), (参照 2020-05-24).  
また2018年3月には慶應義塾大学において「公開ワークショップ 図書館はオープンガバメントに貢献できるか?—行政情報提供と行政支援—」が開催されている。  
“公開ワークショップ(実施報告) 図書館はオープンガバメントに貢献できるか?—行政情報提供と行政支援—”. 慶應義塾大学. [http://user.keio.ac.jp/~lis\\_m/houkoku.html](http://user.keio.ac.jp/~lis_m/houkoku.html), (参照 2020-05-24).

(6) 勢い、本稿では、2019年12月段階での国際図書館連盟(IFLA)による「不確実性や複雑さを増す時代であるからこそ、図書館や情報専門家たる図書館人が提供する情報が重要になってくる」という提起に響きあう内容となっている(E2246参照)。IFLA Headquarters. IFLA TREND REPORT 2019 UPDATE. [https://trends.ifla.org/files/trends/assets/documents/ifla\\_trend\\_report\\_2019.pdf](https://trends.ifla.org/files/trends/assets/documents/ifla_trend_report_2019.pdf), (accessed 2020-05-24).

(7) 佐藤正広. 明治前期における統計調査の調査環境と地方行政. 総務省統計研修所, 2013, 39p., (ディスカッションペーパー, 1). <https://www.stat.go.jp/training/2kenkyu/pdf/mstatou01.pdf>, (参照 2020-05-24).  
佐藤正広編. 近代日本統計史. 晃洋書房, 2020, 291p. など。  
なおこれらの研究については高島正憲氏(関西学院大学経済学部専任講師)の教示を得た。

(8) 佐藤正広. “両大戦間期における公的統計の信頼性—統計編成業務の諸問題とデータの精度について—”. 近代日本統計史. 佐藤正広編. 晃洋書房, 2020, p. 267-285.  
なお、これらは、近代の行政文書を、府県・行政村・大字の各レベルで扱っている実感とある意味符合するところである。例えば、  
平野明夫. “千葉県庁に伝来した文書の謎”. 日本の歴史を解きほぐす. 文学通信, 地方史研究協議会編, 2020, p. 129-144.  
では、明治新政府の度重なる方針転換に振り回される地方機関の様子が描写されている。

(9) 福島幸宏. 占領軍のだまし方. 京都府立総合資料館メールマガジン, 2008, (42). <https://www.prefkyoto.jp/kaidai/maga-ghtml>, (参照 2020-05-24).

(10) “日本の図書館”. 諸外国の公共図書館に関する調査報告書. シー・ディー・アイ, 2005., (平成16年度文部科学省委託事業 図書館の情報拠点化に関する調査研究). p. 273. [https://www.mext.go.jp/a\\_menu/shougai/tosho/houkoku/06082211.htm](https://www.mext.go.jp/a_menu/shougai/tosho/houkoku/06082211.htm), (参照 2020-05-24).

(11) このことは現場の図書館員にも以前から認識されている。例えば、  
田井郁久雄. 開館時間の延長は効果があったか:一地区図書館の事例研究. 図書館界, 2001, 53(2), p. 56-75. [https://doi.org/10.20628/toshokankai.53.2\\_56](https://doi.org/10.20628/toshokankai.53.2_56), (参照 2020-05-24).  
には、入館カウンターの数字をどのように取り扱うかの記述がある。

(12) 例えば、  
原田隆史. 図書館の評価. 図書館界, 2019, 71(2), p. 79. [https://doi.org/10.20628/toshokankai.71.2\\_79](https://doi.org/10.20628/toshokankai.71.2_79), (参照 2020-05-24).  
また、『情報の科学と技術』では、2019年3月に「特集:図書館利用者をデータで把握する」を組んでいる。

[https://www.jstage.jst.go.jp/browse/jkg/69/3/\\_contents/char/ja](https://www.jstage.jst.go.jp/browse/jkg/69/3/_contents/char/ja), (参照 2020-05-24).

(13) 月尾嘉男. デジタルアーカイブ推進コンソーシアム(DAPCON) 2018年度デジタルアーカイブ産業賞受賞講演「デジタルアーカイブの危機」. デジタルアーカイブ学会誌, 2019, 3(4), p. 405-408. [https://doi.org/10.24506/jdsa.3.4\\_405](https://doi.org/10.24506/jdsa.3.4_405), (参照 2020-05-24).

(14) 鷹野凌. “情報の洪水と捏造の時代におけるデジタルアーカイブの意義”. HON.jp News Blog, 2019-08-31. <https://hon.jp/news/1.0/0/26097>, (参照 2020-05-24).

(15) この点、以下のシンポジウムで論じた議論とも連動する。  
第67回日本図書館情報学会研究大会シンポジウム記録:「デジタルアーカイブと図書館」. 日本図書館情報学会誌, 2020, 66(1), p. 41-52.

(16) そして公的記録が、公的には手続きに沿って的確に破棄されているとされているためでもある。この背景については、  
久保亨, 瀬畑源. 国家と秘密:隠される公文書. 集英社, 2014, 206p., (集英社新書, 0759).  
新藤宗幸. 官僚制と公文書:改竄、捏造、忖度の背景. 筑摩書房, 2019, 250p., (ちくま新書, 1407).  
などを参照のこと。そのうえで、ユネスコ(UNESCO)と電子情報保存連合(DPC)が組織内でデジタル保存の重要性を説明するために作成したガイドラインが今後ますます重要となろう(E2200参照)。“Executive Guide on Digital Preservation”. DPC. <https://www.dpconline.org/our-work/dpeg-home>, (accessed 2020-05-24).

(17) 厚生労働省自身が「国内事例については、令和2年5月8日公表分から、データソースを従来の厚生労働省が把握した個票を積み上げたものから、各自治体がウェブサイトで公表している数等を積み上げたものに変更した」とするなど、各種の公的統計に含まれる数字をそのまま直には比較することはすでに困難となっている。  
“新型コロナウイルス感染症について 国内の発生状況”. 厚生労働省. [https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000164708\\_00001.html#kokunaihassei](https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000164708_00001.html#kokunaihassei), (参照 2020-05-24).

(18) 速水融. 日本を襲ったスペイン・インフルエンザ: 人類とウイルスの第一次世界戦争. 藤原書店, 2006, 474p.  
の特に第6章「統計の語るインフルエンザの猖獗」を参照。  
また、  
池田一夫, 藤谷和正, 瀧岡陽子, 神谷信行, 広門雅子, 柳川義勢. 日本におけるスペインかぜの精密分析. 東京都健康安全研究センター年報, 2005, (56), p. 369-374. <http://www.tokyo-eiken.go.jp/sage/sage2005/>, (参照 2020-05-24).  
でも分析が行われている。

(19) われわれは、国レベルの決裁文書自体が「改ざん」されるという、およそ想定外の事態に直面している。これに対応するために作成過程のデータを遺す仕組みの整備等、情報に関わる研究者や専門職団体での議論が必要であろう。  
財務省. 決裁文書の改ざん等に関する調査報告書について. 2018-06-04. [https://www.mof.go.jp/public\\_relations/statement/other/20180604chousahoukoku.html](https://www.mof.go.jp/public_relations/statement/other/20180604chousahoukoku.html), (参照 2020-05-24).

(20) 山本順一. “図書館の社会的責任と知的自由の保障”. 図書館概論. ミネルヴァ書房, 2015, p. 166., (講座・図書館情報学, 2).

[受理: 2020-05-25]

Fukushima Yukihiko

Saving Alternative Information: Thinking about the Future of Libraries from the Problem of Incorrect Statistics

補記: 本稿脱稿後、

佐藤正広. 「統計不信問題」を考える—歴史的視点からの試論. 東京外国語大学国際日本学研究, 2020, プレ創刊号, p. 2-21. <http://repository.tufs.ac.jp/handle/10108/94463>, (参照 2020-06-08). に接した。あわせて参照されたい。



本著作(CA1975)はクリエイティブ・コモンズ 表示 4.0 国際 パブリック・ライセンスの下に提供されています。ライセンスの内容を知りたい方は<https://creativecommons.org/licenses/by/4.0/legalcode.ja>でご確認ください。