

令和5年度
教育データ連携の実証調査研究
(実証Ⅵ)
成果報告書

令和6年3月22日

< 目 次 >

1 背景と目的	1
1.1 背景	1
1.2 目的	1
2 概要	2
2.1 実施対象	2
2.2 実施スケジュール	2
3 実証	3
3.1 データ項目の調査（学習 e ポータル側のデータ調査）	3
3.1.1 出力データ項目	3
3.1.2 各データ項目の出力状況と運用上の取り扱い	3
3.1.3 今後の出力データ項目の拡張	4
3.2 名寄せ手法の検討、検証	5
3.2.1 「学校」の特定	5
3.2.2 「個人」の特定	5
3.2.3 名寄せツール（富士ソフト社提供）の検証	12
3.2.4 名寄せの手法	13
3.3 名寄せ（照合）作業の実証結果	16
3.3.1 作業工数の予実績	16
3.3.2 名寄せ対象の総数	16
3.3.3 名寄せの結果数	17
3.3.4 名寄せできなかったデータと主な原因	17
3.3.5 今後の課題	17
4 まとめ	19
4.1 今後の活用にむけて	19
4.1.1 名寄せに必要な条件	19
4.1.2 学習 e ポータルのデータ管理と名寄せの手法について	19

1 背景と目的

1.1 背景

現在の多くの各自治体において、利用中の校務支援システムと学習 e ポータルは、それぞれ独立したシステムとして導入されており、同じ児童生徒であってもそれぞれのシステム内で名簿や ID を管理しているという状況にある。名簿管理が分かれていると児童生徒情報の登録や更新管理についてはシステムごとに実施しなければならず、年度更新作業もそれぞれで実施が必要となる。また、ID が連携していないことにより相互のデータ利活用もできないため、これらのシステムを効率的かつ有効に利用するためには、名簿連携を行えるよう改善する必要がある。

国の事業としても「教育データ利活用ロードマップ」（令和 4 年 1 月 7 日策定・公表）を踏まえ、教育データの連携を推進するために令和 4 年度にデジタル庁による実証調査研究「教育関連データのデータ連携の実現に向けた実証調査研究」が行われ、この実証研究において国際技術標準規格の「OneRoster」を用いた名簿連携の機能が、参加した各社で実装された。

これによって校務支援システムと学習 e ポータル間で機能上は児童生徒の名簿連携を行うことができるようになったが、現状の運用においては名簿作成の起点である校務支援システムが「UUID」を生成する前に、既に学習 e ポータルの方で MEXCBT 利用の関係上「UUID」を生成し、保持しているという実態がある。学習 e ポータル上の「UUID」を継続利用するためには、学習 e ポータルと校務支援システム間とで名寄せを行い、校務支援システム上の「UUID」を学習 e ポータルの「UUID」で書き戻し、統一する必要がある。

1.2 目的

- ①既に学習 e ポータルで生成、保持している「UUID」がある場合、これを継続利用できるようにするとともに、既存の「UUID」を用いて校務支援システムと学習 e ポータル間で「OneRoster」を用いた名簿連携ができるよう、両システム間の児童生徒データの名寄せを行う。
- ②名寄せ作業で用いた手法、及び課題や留意事項をまとめ、今後、同様に発生するシステム間の名寄せ作業に活用できるようベストプラクティスを導き出す。

2 概要

2.1 実施対象

本実証は、以下を対象として実施した。

○校務支援システム

○学習 e ポータル

○自治体

公立小学校：12 校（児童生徒数：6,864 名）

公立中学校：6 校（児童生徒数：3,252 名）

公立小中一貫校：2 校（児童生徒数：2,695 名）

2.2 実施スケジュール

本実証は、以下のスケジュールの通り実施した。

表 2-2-1 スケジュール

工程	時期
1. データ項目の調査	
学習 e ポータル側のデータ調査	2024 年 2 月上旬
2. 名寄せ手法の検討、検証	
名寄せツール（富士ソフト社提供）の検証	2024 年 2 月中旬～下旬
名寄せ手法の検討、検証	2024 年 2 月中旬～下旬
独自ツールの作成	2024 年 2 月下旬～3 月上旬
3. 名寄せ（照合）作業の実証	
実データを用いた名寄せ（照合）作業	2024 年 2 月下旬～3 月中旬
4. 実証結果のとりまとめ	
課題、留意事項等の整理・ベストプラクティスの取りまとめ	2024 年 3 月中旬～下旬
報告	2024 年 3 月下旬

3 実証

3.1 データ項目の調査（学習 e ポータル側のデータ調査）

3.1.1 出力データ項目

本実証において名寄せに使用した学習 e ポータルからの出力データ項目は以下の通り。

表 3-1-1 学習eポータルからの出力データ項目

No	データ項目	データ型	Not Null	説明
1	表示名	文字列	No	自由入力。姓名もしくは出席番号が入る可能性がある。
2	ユーザーID	文字列	Yes	テナントごとにユニークになる学習 e ポータルのシステム内 ID
3	学校	文字列	Yes	学校名に相当するが、運用上実名でない場合あり。
4	学年	文字列	Yes	「小学 1 年」～「小学 6 年」、「中学 1 年」～「中学 3 年」が出力される。
5	クラス	文字列	Yes	クラス名に「組」が付加した形式で出力される。
6	識別 ID	文字列	No	MEXCBT で使う UUID を出力。

また、名寄せの実施にあたり有効だと考えられる以下のデータ項目については、本実証で対象とした学習 e ポータルにおいては、システムとしてデータ項目を保持しない仕組みとなっていたため、データ自体が出力不可であった。

表 3-1-2 学習eポータルから出力不可能なデータ項目

No	データ項目	データ型	備考
1	出席番号	数値	出力不可。「表示名」に設定される可能性もあり。
2	姓(フリガナ)	文字列	出力不可。「表示名」に設定される可能性もあり。
3	名(フリガナ)	文字列	出力不可。「表示名」に設定される可能性もあり。
4	生年月日	日付	出力不可。
5	性別	文字列	出力不可。

3.1.2 各データ項目の出力状況と運用上の取り扱い

本実証における学習 e ポータルから出力可能な各データ項目の出力状況と運用上の取り扱いについて、以下の通り確認を行った。

No.1 表示名

- ・教育委員会からは「表示名」には出席番号等、個人の特定に繋がる情報は含めないように各校にお願いしているが、学校での運用上は「表示名」に個人を識別可能な情報（出席番号など）を設定して、識別している。
- ・運用方法は各学校に委任しており、学校単位で運用ルールを決めているが、教員が独自に変更している学校もある。
- ・児童生徒が独自に「表示名」を変更している可能性もある。

No.2 ユーザーID

- ・システムが付与する ID で出力規則は統一されており、ユーザーが変更することはない。
- ・年組番と関連性はない。

No. 3 学校

- ・本実証では、「001 小学校」、「101 中学校」、「001101 小中一貫校」といった形式で学校名の特定はできないようになっている。
- ・出力規則は統一されて出力されており、ユーザーが変更することはない。

No. 4 学年

- ・システムの命名規則に沿って「小学 1 年」～「小学 6 年」、「中学 1 年」～「中学 3 年」が出力される。
- ・出力規則は統一されて出力されており、ユーザーが変更することはない。

No. 5 クラス

- ・クラス名に「組」が付加した形式で出力される。
- ・出力規則は統一されて出力されており、ユーザーが変更することはない。
- ・年度更新時に一時的に児童生徒アカウントを格納する実在しないクラスもデータとして存在する。

No. 6 識別 ID

- ・学習 e ポータル標準モデルで規定されている UUID が出力される。
- ・出力規則は統一されて出力されており、ユーザーが変更することはない。

3.1.3 今後の出力データ項目の拡張

本実証にて使用した学習 e ポータルについては、実証中にデータエクスポート機能の実装が進められており、本実証に利用することはできなかったが、今後は以下の項目が出力可能となり、今後の名寄せに活用できる可能性がある。

表 3-1-3 学習eポータルからの出力データ項目（追加）

No	データ項目	データ型	Not Null	説明
1	外部認証用メールアドレス	文字列	No	必須ではないため、値なし場合もあり。 SSO 用のメールアドレス (Google or MS) を出力。
2	年度	文字列	Yes	
3	ユーザー種別	文字列	Yes	「児童生徒」、「一般教員」、「情報担当者」、「管理職教員」が入る。

3.2 名寄せ手法の検討、検証

3.2.1 「学校」の特定

「3.1 データ項目の調査（学習 e ポータル側のデータ調査）」に記載の通り、学校を識別するデータ項目「学校」には運用上の実名が使われておらず、「001 小学校」、「101 中学校」、「001101 小中一貫校」といった形式で出力されているため、まずは学校の特定が必要となった。学習 e ポータル側の学校の並び順は基本的に文科省の学校コードの並び順であるが、確認のため以下の要素で判断を行った。

I. 学年ごとのクラス名とクラス数の一致確認

学習 e ポータル側から出力された学校の各学年のクラス名とクラス数が校務支援システム側の登録内容と一致するか確認を行った。

II. 「表示名」に個人名が含まれているデータを利用

おそらく児童生徒が独自に「表示名」を変更しているケースだと考えられるが、学習 e ポータル側の「表示名」に個人名を登録しているケースがあり、各校においては少なくとも1件はそのようなデータが存在していた。このデータを利用して、該当する学校の該当クラスにその生徒が本当に存在するか校務支援システム側のデータと照合を行い、確認を行った。

以上の方法から学習 e ポータル側の学校の並び順は、文科省の学校コードの並び順と一致していることの確認が取れた。

3.2.2 「個人」の特定

「表 3-1-1 学習 e ポータルからの出力データ項目」に示した通り、学校、学年、クラスについては必須項目として学習 e ポータルの DB にデータ項目の設定があるため明確に識別することが可能であるが、出席番号や名前など、個人が明確に識別できるデータ項目がない。

基本的に学習 e ポータル上では、個人を識別できるためのデータ項目の設定がないが、「3.1.2 各データ項目の出力状況と運用上の取り扱い」にも示した通り、学校での運用上は「表示名」に個人を識別可能な情報（出席番号など）を設定して識別していることから、姓名もしくは出席番号の入る可能性のある「表示名」で個人の識別を試みることにした。

まずは、「表示名」の命名規則について、どのようになっているか全データを通して以下の通り確認を行った。

I. 「表示名」の命名規則の確認

- ・基本的には学校で統一された命名規則になっているが、学年単位、もしくはクラス単位で命名規則を変えて設定しているケースがある。
- ・児童生徒本人が「表示名」を変えているケースがあり、同一学校、学年、クラス内であっても個人によって全く命名規則が異なっているケースがある。
- ・「表示名」の命名規則として出現したパターンは以下の通りである。

表 3-2-2 「表示名」の命名規則パターン

※赤文字:個人識別が可能となるデータ

識別方法	パターン No	命名規則	詳細	個人識別 可否
出席番号	1-1	出席番号	例)1 →“1”は出席番号 1 番を表す	可
	1-2	出席番号 +何らかの文字列	例)01A あ →“01”は出席番号 1 番を表す “A あ”は何らかの文字(何を表しているかは不明)	可
	1-3	年 +組 +出席番号	例)3-2-1 →“3”は 3 年を表す “2”は 2 組を表す “1”は出席番号 1 番を表す	可
	1-4	年 +組 +出席番号	例)3201 →“3”は 3 年を表す “2”は 2 組を表す “01”は出席番号 1 番を表す	可
	1-5	何らかの番号 +年 +組 +出席番号	例)103201 →“10”は何らかの番号(何を表しているかは不明) “32”は 3 年 2 組を表す “01”は出席番号 1 番を表す	可
	1-6	何らかの文字列 +年 +組 +出席番号	例)s32@01 →“s”は何らかの文字列(何を表しているかは不明) “32”は 3 年 2 組を表す “01”は出席番号 1 番を表す	可
	1-7	何らかの文字列 +年 +組 +出席番号 +何らかの番号	例)s32@01987 →“s”は何らかの文字列(何を表しているかは不明) “32”は 3 年 2 組を表す “01”は出席番号 1 番を表す “987”は何らかの番号(何を表しているかは不明)	可
	1-8	何らかの文字列 +何らかの番号 +出席番号	例)あ 0901 →“あ”は何らかの文字列(何を表しているかは不明) “09”は何らかの番号(何を表しているかは不明) “01”は出席番号 1 番を表す	可
	1-9	何らかの番号 +組 +出席番号	例)203201 →“203”は何らかの番号(何を表しているかは不明) “2”は 2 組を表す “01”は出席番号 1 番を表す	可
	1-10	+年 +組 +出席番号	例)3201 →“3”は 3 年を表す “2”は 2 組を表す “01”は出席番号 1 番を表す	可
	1-11	+年 +組 +出席番号	例)3 年 2 組 1 番 →“3 年 2 組”は 3 年 2 組を表す “1 番”は出席番号 1 番を表す	可
	1-12	+出席番号 +年 +組	例)010302 →“01”は出席番号 1 番を表す “0302”は 3 年 2 組を表す	可
	1-13	(出席番号) +ユーザーID の一部	例)(01) 45678 →“(01)”は出席番号 1 番を表す “45678”はユーザーID の一部を表す	可
	1-14	ユーザーID の一部 +(出席番号) (※ユーザーID の一部と出席番号の間にスペースが入るケースもあり)	例)45678(1)、45678 (1) →“45678”はユーザーID の一部を表す “(1)”は出席番号 1 番を表す	可
出席番号	1-15	ユーザーID の一部 +(出席番号)	例)45678(1番) →“45678”はユーザーID の一部を表す “(1番)”は出席番号 1 番を表す	可

識別方法	パターン No	命名規則	詳細	個人識別 可否
	1-16	ユーザーID の一部 +＜出席番号＞	例) 45678<1> →“45678”はユーザーID の一部を表す “<1>”は出席番号 1 番を表す	可
	1-17	ユーザーID +出席番号	例) 123-4567801 →“123-45678”はユーザーID を表す “01”は出席番号 1 番を表す	可
	1-18	何らかの番号 +年 +組 +出席番号 +絵文字、キーワード等	例) 103201☺、 103201 ^-^、 103201 BOOK、 103201 です。。。. →“10”は何らかの番号(何を表しているかは不明) “32”は 3 年 2 組を表す “01”は出席番号 1 番を表す “☺”、“^-^”、“BOOK”、“です。。。.”は 意味を持たない文字列	可
出席番号と 個人名	2-1	+年 +組 +出席番号 +姓 +名 (※年組と出席番号と 姓名の間にスペース が入るケースもあり)	例) 3 年 2 組 1 番 東書花子、3 年 2 組 1 番 東書花子 →“3 年 2 組”は 3 年 2 組を表す “1 番”は出席番号 1 番を表す “東書”は姓を表す “花子”は名を表す	可
	2-2	年 +組 +出席番号 +姓 +名 (※出席番号、姓名 の間にスペースが入 るケースもあり)	例) 3-2-1 東書花子、3-2-1 東書 花子、 3-2-1 東書花子、3-2-1 東書 花子 3-2 1 東書花子、3-2 1 東書 花子 3-2 1 番 東書花子、3-2 1 番 東書 花子 →“3”は 3 年を表す “2”は 2 組を表す “1”は出席番号 1 番を表す “東書”は姓を表す “花子”は名を表す	可
	2-3	年 +組 +出席番号 +姓ふりがな +名ふりがな	例) 3-2-1 とうしょはなこ →“3”は 3 年を表す “2”は 2 組を表す “1”は出席番号 1 番を表す “とうしょ”は姓のふりがなを表す “はなこ”は名のふりがなを表す	可
	2-4	年 +組 +出席番号 +姓 (※出席番号、姓の 間にスペースが入る ケースもあり)	例) 3-2-1 東書、3-2-1 東書 →“3”は 3 年を表す “2”は 2 組を表す “1”は出席番号 1 番を表す “東書”は姓を表す	可
	2-5	年 +組 +出席番号 +姓 (※出席番号、姓の 間にスペースが入る ケースもあり)	例) 3-2-1 花子、3-2-1 花子 →“3”は 3 年を表す “2”は 2 組を表す “1”は出席番号 1 番を表す “花子”は名を表す	可
出席番号と 個人名	2-6	年 +組 +出席番号 +名ふりがな	例) 3-2-1 はなこ →“3”は 3 年を表す “2”は 2 組を表す “1”は出席番号 1 番を表す “はなこ”は名のふりがなを表す	可

識別方法	パターン No	命名規則	詳細	個人識別 可否
	<u>2-7</u>	年 +組 +出席番号 +姓ローマ字表記 +名ローマ字表記	例) 3-2-1 tosho hanako →“3”は3年を表す “2”は2組を表す “1”は出席番号1番を表す “tosho”は姓のローマ字表記を表す “hanako”は名のローマ字表記を表す	可
	<u>2-8</u>	年 +組 +出席番号 +名ローマ字表記	例) 3-2-1 Hanako →“3”は3年を表す “2”は2組を表す “1”は出席番号1番を表す “Hanako”は名のローマ字表記を表す	可
	<u>2-9</u>	何らかの番号 +年 +組 +出席番号 +姓 +名 (※出席番号と姓名 の間にスペースやア ンダーバーが入るケ ースもあり)	例) 103201 東書花子、103201 東書花子、 103201_東書花子 →“10”は何らかの番号(何を表しているかは不明) “32”は3年2組を表す “01”は出席番号1番を表す “東書”は姓を表す “花子”は名を表す	可
	<u>2-10</u>	何らかの番号 +年 +組 +出席番号 +姓	例) 103201 東書 →“10”は何らかの番号(何を表しているかは不明) “32”は3年2組を表す “01”は出席番号1番を表す “東書”は姓を表す	可
	<u>2-11</u>	何らかの番号 +年 +組 +出席番号 +名	例) 103201 花子 →“10”は何らかの番号(何を表しているかは不明) “32”は3年2組を表す “01”は出席番号1番を表す “花子”は名を表す	可
	<u>2-12</u>	+出席番号 +姓 +名 (※出席番号と姓名 の間にピリオドが入 るケースもあり)	例) 01 東書花子、01_東書花子 →“01”は出席番号1番を表す “東書”は姓を表す “花子”は名を表す	可
	<u>2-13</u>	何らかの番号 +組 +出席番号 +姓 +名 (※出席番号と姓名 の間にスペースやア ンダーバーが入るケ ースもあり)	例) 103201 東書花子、103201 東書花子、 103201_東書花子 →“10”は何らかの番号(何を表しているかは不明) “32”は3年2組を表す “01”は出席番号1番を表す “東書”は姓を表す “花子”は名を表す	可
	<u>2-14</u>	+出席番号 +年 +組 +姓ふりがな +名ふりがな	例) 010302 とうしょはなこ →“01”は出席番号1番を表す “0302”は3年2組を表す “とうしょ”は姓のふりがなを表す “はなこ”は名のふりがなを表す	可

識別方法	パターン No	命名規則	詳細	個人識別 可否
出席番号と 個人名	<u>2-15</u>	姓 + 名 + 何らかの番号 + 年 + 組 + 出席番号 (※姓名が特殊な括弧で囲まれるケースもあり)	例) 東書花子 53201、【東書花子】53201 →“東書”は姓を表す “花子”は名を表す “5”は何らかの番号(何を表しているかは不明) “32”は 3 年 2 組を表す “01”は出席番号 1 番を表す	可
	<u>2-16</u>	姓ふりがな + 年 + 組 + 出席番号	例) とうしょ 3201 →“とうしょ”は姓のふりがなを表す “32”は 3 年 2 組を表す “01”は出席番号 1 番を表す	可
	<u>2-17</u>	何らかの番号 + 年 + 組 + 出席番号 + 名ふりがな	例) 103201 はなこ →“10”は何らかの番号(何を表しているかは不明) “32”は 3 年 2 組を表す “01”は出席番号 1 番を表す “はなこ”は名のふりがなを表す	可
	<u>2-18</u>	姓ふりがな + 名ふりがな + 何らかの番号 + 年 + 組 + 出席番号	例) とうしょ はなこ 103201 →“とうしょ”は姓のふりがなを表す “はなこ”は名のふりがなを表す “10”は何らかの番号(何を表しているかは不明) “32”は 3 年 2 組を表す “01”は出席番号 1 番を表す	可
	<u>2-19</u>	何らかの番号 + 年 + 組 + 出席番号 + 名の一部	例) 103201 はな →“10”は何らかの番号(何を表しているかは不明) “32”は 3 年 2 組を表す “01”は出席番号 1 番を表す “はな”は名のふりがなの一部を表す	可
	<u>2-20</u>	姓 + 名 + 出席番号	例) 東書 花子 1 番 →“東書”は姓を表す “花子”は名を表す “1 番”は出席番号 1 番を表す	可
	<u>2-21</u>	姓ふりがな + 名ふりがな + (出席番号)	例) とうしょ はなこ(1) →“とうしょ”は姓のふりがなを表す “はなこ”は名のふりがなを表す “(1)”は出席番号 1 番を表す	可
	<u>2-22</u>	ユーザーID の一部 + (出席番号) + 姓	例) 45678(1)東書 →“45678”はユーザーID の一部を表す “(1)”は出席番号 1 番を表す “東書”は姓を表す	可
	<u>2-23</u>	姓 + ユーザーID の一部 + 出席番号	例) 東書 45678-1 →“東書”は姓を表す “45678”はユーザーID の一部を表す “-1”は出席番号 1 番を表す	可
	<u>2-24</u>	出席番号 + 名ふりがな	例) 01 はなこ →“01”は出席番号 1 番を表す “はなこ”は名ふりがなを表す	可
	<u>2-25</u>	ユーザーID の一部 + (出席番号)	例) 1 45678(01) →“45678”はユーザーID の一部を表す “(01)”は出席番号 1 番を表す	可
	<u>2-26</u>	ユーザーID の一部 + 出席番号	例) 1 45678 1 番 →“45678”はユーザーID の一部を表す “1 番”は出席番号 1 番を表す	可
	<u>2-27</u>	(出席番号) + ユーザーID	例) (01)123-45678 →“(01)”は出席番号 1 番を表す “123-45678”はユーザーID を表す	可

識別方法	パターン No	命名規則	詳細	個人識別 可否
出席番号と 個人名	<u>2-28</u>	出席番号 +ユーザーID (※出席番号とユーザーIDの間にハイフンが入るケースもあり)	例) 1 45678、1-45678 →“1”は出席番号 1 を表す “45678”はユーザーID の一部を表す	可
出席番号と 個人名のイ ニシャル	<u>3-1</u>	何らかの番号 +年 +組 +出席番号 +姓名のイニシャル	例) 103201 H・T、103201 TH、 →“10”は何らかの番号(何を表しているかは不明) “32”は 3 年 2 組を表す “01”は出席番号 1 番を表す “H・T”、“TH”は名イニシャルと姓イニシャルを表す	可
	<u>3-2</u>	出席番号 +年 +組 +姓名のイニシャル	例) 010302th →“01”は出席番号 1 番を表す “0302”は 3 年 2 組を表す “th”は姓イニシャルと名イニシャルを表す	可
	<u>3-3</u>	出席番号 +姓名のイニシャル	例) 01 T.H、01 TO.H、01 TSH.H →“01”は出席番号 1 番を表す “T.H”、“TO・H”は姓イニシャルと名イニシャルを表す	可
	<u>3-4</u>	何らかの番号 +年 +組 +出席番号 +名のイニシャル	例) 103201h、h103201 →“10”は何らかの番号(何を表しているかは不明) “32”は 3 年 2 組を表す “01”は出席番号 1 番を表す “h”は名イニシャルを表す	可
	<u>3-5</u>	何らかの番号 +姓名のイニシャル +出席番号	例) 12TH01 →“12”は何らかの番号(何を表しているかは不明) “TH”は姓イニシャルと名イニシャルを表す “01”は出席番号 1 番を表す	可
出席番号と 個人名のローマ字表記	<u>4-1</u>	何らかの番号 +年 +組 +出席番号 +名のローマ字表記	例) 103201 hanako →“10”は何らかの番号(何を表しているかは不明) “32”は 3 年 2 組を表す “01”は出席番号 1 番を表す “hanako”は名のローマ字表記を表す	可
個人名	<u>5-1</u>	年 +組 +姓 +名 (※姓名の間にスペースが入るケースもあり)	例) 3-2 東書花子、3-2 東書 花子 →“3-2”は 3 年 2 組を表す “東書”は姓を表す “花子”は名を表す	可
	<u>5-2</u>	年 +組 +姓 +名 (※姓名の間にスペースが入るケースもあり)	例) 3 年 2 組東書花子 →“3 年 2 組”は 3 年 2 組を表す “東書”は姓を表す “花子”は名を表す	可
	<u>5-3</u>	姓 +名 (※姓名の間にスペースが入るケースもあり)	例) 東書花子、東書 花子 →“東書”は姓を表す “花子”は名を表す	可

識別方法	パターン No	命名規則	詳細	個人識別 可否
個人名	5-4	姓ふりがな + 名ふりがな (※姓ふりがなと名ふりがなの間にスペースが入るケース、特殊な括弧で囲まれるケースもあり)	例)とうしょはなこ、とうしょ はなこ、【とうしょはなこ】 →“とうしょ”は姓のふりがなを表す “はなこ”は名のふりがなを表す	可
	5-5	姓 + 名ふりがな	例)東書はなこ →“東書”は姓を表す “はなこ”は名のふりがなを表す	可
	5-6	何らかの番号 + (姓 + 名)	例)p99-0123(東書花子) →“p99-0123”は何らかの番号(何を表しているかは不明) “東書花子”は姓名を表す	可
	5-7	姓 + 名のローマ字表記	例)東書 HANAKO →“東書”は姓を表す “HANAKO”は名のローマ字表記を表す	可
	5-8	姓	例)東書 →“東書”は姓を表す	否
	5-9	名	例)花子 →“花子”は名を表す	否
	5-10	姓ふりがな	例)とうしょ →“とうしょ”は姓のふりがなを表す	否
	5-11	名ふりがな	例)はなこ →“はなこ”は名のふりがなを表す	否
	5-12	名 + 年 + 組	例)花子 3-2 →“花子”は名を表す “3-2”は 3 年 2 組を表す	否
	5-13	姓ふりがな + 名の一部 (※姓名の間にスペースが入るケースもあり)	例)とうしょ 花 →“とうしょ”は姓ふりがなを表す “花”は名の一部を表す	可
	5-14	名の一部	例)ハナ →“ハナ”は名の一部を表す	否
	5-15	名前 + ユーザーID	例)花子 123-45678 →“花子”は名を表す “123-45678”はユーザーIDを表す	否
個人名のローマ字表記	6-1	姓 + 名のローマ字表記	例)hanako_tosho、Hanako Tosho →“hanako_tosho”、“Hanako Tosho”は姓のローマ字表記と名のローマ字表記を表す	可
	6-2	姓のローマ字表記	例)tosho →“tosho”は姓のローマ字表記を表す	否
	6-3	名のローマ字表記	例)hanako →“hanako”は名のローマ字表記を表す	否
	6-4	名のローマ字表記 + 年 + 組	例)hanako3-2 →“hanako”は名のローマ字表記を表す “3-2”は 3 年 2 組を表す	否
個人名のイニシャル	7-1	姓名のイニシャル	例)T:H、TH →“T:H”、“TH”は姓イニシャルと名イニシャルを表す	否
	7-2	名のイニシャル + 組 + 名のローマ字の一部	例)h2 組 nako →“h”は名イニシャルを表す “2 組”は 2 組を表す “nako”は名のローマ字の一部を表す	否
ユーザーID (※個人識別できず)	8-1	ユーザーID	例)123-45678 →“123-45678”はユーザ ID を表す	否

識別方法	パターン No	命名規則	詳細	個人識別 可否
その他	9-1	日本語、英語以外の 表記	例) 공공	否
	9-2	転出データであることを 知らせる標記	例) 転出、01 転、転 01、転 1001	否
	9-3	予備データであることを 知らせる標記	例) 予備アカウント	否
	9-4	意味を持たない文字 列	例) エラーのためこのページはご覧になれません	否

3.2.3 名寄せツール（富士ソフト社提供）の検証

本実証の取りまとめ会社である富士ソフト社から学習 e ポータルから出力したデータと校務支援システムから出力したデータを比較し、名寄せを行う Excel マクロを用いた名寄せツールが提供されたため、これが利用可能か検証を行った。

I. 出力データのファイル形式

名寄せツールでは学習 e ポータル、及び校務支援システムから名寄せに必要なデータが出力された excel 形式または csv 形式ファイルが必要となるが、本実証で対象とした校務支援システムと学習 e ポータルにおいてはそれぞれ以下のように用意が可能であった。

- ・校務支援システム：自治体内の全学校のデータが一括で出力されたファイル（CSV 形式）
- ・学習 e ポータル：学校ごとのデータが分割して出力されるファイル（CSV 形式）

II. データ項目

名寄せツールでは、校務支援システムと学習 e ポータルとで、どのデータ項目を比較するか事前に設定が必要となるが、名寄せに必要な項目が揃っているか両システムから出力されたファイルを確認したところ以下の通りであった。

表 3-2-3 校務支援システムと学習eポータルの主な名寄せデータ項目の比較

※赤文字：個人識別に必要なデータ

No	データ項目	校務支援システム		学習 e ポータル	
		有無	データ形式	有無	データ形式
1	学校	○	学校正式名称を文字列形式で保持	△	「001 小学校」、「101 中学校」、 「001101 小中一貫校」等のコードで保持
2	学年	○	コード(半角数字 2 桁)で保持 02: 小学 1 年生～10: 中学 3 年生	○	「小学 1 年」～「小学 6 年」、 「中学 1 年」～「中学 3 年」という 文字列形式で保持
3	クラス	○	クラス名を文字列形式で保持	○	クラス名に「組」が付加した文字列形式 で保持
4	出席番号	○	半角数字 1 桁～3 桁の数値で保持	×	—
5	姓	○	全半角の文字列形式で保持	×	—
6	名	○	全半角の文字列形式で保持	×	—
7	姓ふりがな	○	ひらがな、カタカナ全角、カタカナ半角 のいずれかで保持(※統一なし)	×	—
8	名ふりがな	○	ひらがな、カタカナ全角、カタカナ半角 のいずれかで保持(※統一なし)	×	—
9	生年月日	○	生まれ年(半角数字 4 桁)、 生まれ月(半角数字 1 桁～2 桁)、 生まれ日(半角数字 1 桁～2 桁) の3項目に分けて保持	×	—

No	データ項目	校務支援システム		学習 e ポータル	
		有無	データ形式	有無	データ形式
10	性別	○	コード(半角数字 1 桁)で保持 1:男、2:女	×	—
11	表示名	×	—	○	全半角の文字列形式で保持
12	ユーザーID	○	システム内の独自コードとして 「0000000」形式で保持	○	システム内の独自コードとして 「000-00000」形式で保持

名寄せを行うには最低限、①出席番号、②姓名または姓名（フリガナ）、③生年月日のいずれかのデータが必要であると考えられるが、本実証で対象とした学習 e ポータルにおいては、システムの仕組みとして、いずれのデータも固定のデータ項目として保持していない。

学校の運用上、「表示名」というデータ項目にルール不統一で上記の①②③のデータ保持する可能性があるが、「表 3-2-2 「表示名」の命名規則パターン」にも記載した通り、自由度の高い設定項目であるため、約 80 通りもの登録方法が存在する。どの部分が①出席番号、②姓名または姓名（フリガナ）、③生年月日を表しているのかについても一見してわからないため、汎用的な名寄せツールを用いてデータ項目の比較と効率的な名寄せをすることは難しいと判断し、別途独自のツールを用いて名寄せを行うこととした。

3.2.4 名寄せの手法

名寄せ作業の方針として、まず独自に作成したツールにより「表示名」にある程度の法則性があるデータに対して機械的に名寄せを行い、次に人力作業によってツールによる名寄せ結果の確認を行い、さらに名寄せできなかったデータの再名寄せを手動で行うこととした。

手順は以下の通り行った。

I. ツールによる名寄せ処理

VBA で作成したツールを用いて Excel 上で以下の流れで名寄せの処理を行った。

※処理説明を簡潔にするため便宜上、以下のように表現する。

A ファイル：校務支援システムから出力したデータファイル

B ファイル：学習 e ポータルから出力したデータファイル

C ファイル：名寄せ結果を出力するファイル

1. 前処理

- 1) A ファイルの情報を元に、学校単位の C ファイルを作成
- 2) A ファイルの情報を元に、学校単位の C ファイル内に 1 クラスに対して 1 シートを作成
- 3) A ファイルから該当する学校、学年、クラスのデータを抽出し、出席番号順にソートして C ファイルの該当クラスのシートに転記

2. 共通処理

- 1) 文字列比較を行う場合、数値、アルファベットは半角、それ以外は全角に揃えて比較
- 2) 文字列比較を行う場合、スペース、ハイフン、アンダーバーは取り除いて比較

3. 名寄せ作業

- 1) 前処理で作成した、クラスのシート毎に以下の処理を繰り返し実施。
- 2) B ファイルから該当する学校、学年、クラスのデータを抽出し、「表示名」でソート

- 3) Bファイルのレコードの「表示名」に数値が含まれている場合、法則性があるか判定（共通文字列部分と数値の連番部分に分けることができるか判定）
→法則性がある場合、連番部分がAファイルから抽出した出席番号と一致する行にBファイルの該当レコードを転記（Cファイルに転記）
- 4) 処理 3) で名寄せできなかったBファイルのレコードについて、「表示名」に数値以外が含まれている場合、Aファイルから抽出した姓、名、姓ふりがな、名ふりがなのうち、姓+名、姓ふりがな+名ふりがな、姓+名ふりがな、姓ふりがな+名のいずれかと一致する文字列が含まれているか判定
→文字列が含まれている場合はAファイルから抽出したレコードと一致する行にBファイルの該当レコードを転記（Cファイルに転記）
- 5) Aファイルから抽出した姓ふりがな、名ふりがなをローマ字に変換したデータ、イニシャルに変換したデータをテンポラリーファイル上に保持。処理 4) で名寄せできなかったBファイルのレコードについて、ローマ字変換したもの、イニシャル変換したものと一致する文字列が含まれているか判定
→文字列が含まれている場合はAファイルから抽出したレコードと一致する行にBファイルの該当レコードを転記（Cファイルに転記）
- 6) 処理 3) ～5) で名寄せできなかったBファイルのレコードは、Aファイルから抽出したレコードの最下行の次の行にまとめて転記（Cファイルに転記）

Ⅱ. 手動による確認と残件の名寄せ処理

「Ⅰ. ツールによる名寄せ処理」で名寄せした結果に誤りがないか目視で確認を行うとともに、名寄せできなかったデータが別の方法で名寄せ可能か確認を行った。

別の方法で名寄せができたパターンは以下の通り。

1. 「表示名」が昨年度の年組番号の内容と一致することによる名寄せ

名寄せできなかったBファイルのレコードのうち「表示名」の命名規則が年組番となっている同クラス内の他レコードと体系的には類似しているが、クラスや出席番号が一致しないものが存在した。

「表示名」の値から見て昨年度の内容のようにも見受けられたため、名寄せできなかったAファイルの児童生徒についても校務支援システム側で履歴データを調べてみた。すると、名寄せできなかったBファイルのレコードの「表示名」が、名寄せできなかったAファイルの児童生徒の昨年度の年組番号と一致することがわかった。

運用上、年度初めに学校にて「表示名」を再設定しているケースが多いが、設定変更から漏れて昨年度に命名された状態のままとなったケースではないかと想定される。

学習eポータル側のデータと校務支援システム側のデータで、名寄せできなかった残データ数がそれぞれ一致しており、名寄せできなかったBファイルのレコードの「表示名」とAファイルの児童生徒の昨年度の年組番号が一致した場合には、該当するデータとして扱い、名寄せを行った。

2. 「表示名」が一部のデータと一致することによる名寄せ

学習eポータル側のデータと校務支援システム側のデータで、名寄せできなかった残データ数がそれぞれ一致しており、学習eポータルの「表示名」が校務支援システム側の保持する姓名の漢字、ひらがな、カタカナ、ローマ字、イニシャルのいずれか

と部分一致し、それ以外に一致するデータが存在しなかった場合には、該当するデータとして扱い、名寄せを行った。

本実証においてはツール作成の期間が十分に確保できなかったことにより、トライアンドエラーを繰り返しながら名寄せを実施するという状況ではあったが、大半はツールを使っての名寄せが有効であった。しかし、ツール自体の品質が結果の正当性に大きな影響を与えるため、いずれにしても最終的な人の手と目による確認が不可欠であった。

3.3 名寄せ（照合）作業の実証結果

3.3.1 作業工数の予実績

本実証における作業工数の予実績は以下の通り。

表 3-3-1 作業工数の予実績

No	工程	作業詳細	工数(人日)	
			予定	実績
1	データ項目の調査	学習 e ポータル側のデータ調査	6	6
2	名寄せ手法の検討、検証	名寄せツール(富士ソフト社提供)の検証	6	2
		名寄せ手法の検討、検証		4
		独自ツールの作成	-	8
3	名寄せ(照合)作業の実証	実データを用いた名寄せ(照合)作業	16	16
		結果の確認、検証	4	4
4	実証結果のとりまとめ	成果報告書の作成	8	5
合計			40	45

計画では、提供される汎用の名寄せツール（富士ソフト社提供）を使用して名寄せを行う予定であったが、実証対象のデータには適用が難しいことが判明し、独自のツールを作成することとした。計画中で工数をおさめられるか調整を試みたが、実データの登録パターンが想定よりも多く存在したため、分析とツールへの反映に時間がかかり全体的なコストは超過してしまった。しかし、独自ツールの作成の段階で、並行して報告書へ記載する内容についてある程度まとめておくことができたため、成果報告書の作成のフェーズを予定より短縮することができ、多少のリカバリーができた。

3.3.2 名寄せ対象の総数

もともと学習 e ポータルから出力したファイル内のレコード数と校務支援システムから出力したファイル内の児童生徒数については差異があり、以下の通り母数が異なる状態であった。

※ただし、学習 e ポータルからの出力ファイルに存在する、“年度更新時に一時的に児童生徒アカウントを格納する実在しないクラス”に関するデータは母数として除いている。

表 3-3-2 名寄せ対象の総レコード数

No	学校種別	校務支援システム (A)	学習 e ポータル (B)	差異 (B-A)
1	小学校	6,864	6,946	82
2	中学校	3,252	3,251	-1
3	小中一貫校	2,695	2,970	275

学習 e ポータルには支援員が使用するデータや予備データ、転出した児童生徒のデータ等が含まれているため、校務支援システムよりも数が多くなると想定されるが、中学校に関しては、学習 e ポータルの方が、総数が少なくなっている。転入生などに対して新たな ID 付与が追いついていないなどのケースも考えられるが、正確な理由についてはさらなる分析が必要である。

3.3.3 名寄せの結果数

本実証で実施した名寄せの結果数と成功率、失敗率は以下の通りである。

表 3-3-3 名寄せ結果数

No	学校種別	名寄せできた数 (C)	名寄せできなかった数 (D)	名寄せの成功率 (C/A) × 100	名寄せの失敗率 (B/A) × 100
1	小学校	6,727	137	98 %	2 %
2	中学校	2,657	595	82 %	18 %
3	小中一貫校	2,653	42	98 %	2 %

小学校、小中一貫校についてはほとんど名寄せができたが、中学校については2割近く名寄せできないケースがあった。

3.3.4 名寄せできなかったデータと主な原因

名寄せできなかった原因とデータ数は以下の通り

表 3-3-4 名寄せできなかった原因別の件数

No	学校種別	①手がかりがない	②候補が2つある
1	小学校	112	25
2	中学校	584	11
3	小中一貫校	37	5

名寄せできなかった原因は、「表 3-3-4 名寄せできなかった原因別の件数」に示す通り、大きく分けて2つあり、「①手がかりがない」に該当するデータは、学習 e ポータルの「表示名」に出席番号や姓名に関わる情報が設定されておらず対応するデータを特定できなかったケースである。もう一つの「②候補が2つある」は「表示名」に出席番号や姓名に関わる情報が設定されているが、対応するデータが複数件あり、一意に特定できなかったケースである。

中学校で名寄せができなかった原因の大半は「①手がかりがない」に該当するが、これはある学校において全校的に学習 e ポータルの「表示名」がデフォルト値から何も変更していない状態（※個人を特定できる情報のない状態）であったことが大きく影響している。「表示名」は、学校単位で運用ルールを決めて運用するとのことであったが、おそらく当該校では、学校として学習 e ポータル自体をあまり利用していない可能性が考えられる。

3.3.5 今後の課題

I. 個人を特定できなかったデータへの対応

本実証では、個人を識別できる可能性の高いデータを用いて名寄せを行ったが、それだけでは名寄せができなかったデータが多く存在した。今後も学習 e ポータル上のアカウントについて個人識別を可能とする明確な情報が提供されない限り、現状保持する限られたデータの中から名寄せを試みなければならないため、どのような項目を用いてどこまで名寄せが可能かについても検証を続ける必要があるといえる。

しかし、不確実なデータを用いて名寄せの確度を高める方法について試行錯誤を続けることにも限界があるため、可能であれば学習 e ポータル上に個人を識別することのできる確実な情報を保持することにした方が望ましいと考えられる。

Ⅱ．名寄せ結果の正当性の確認

本実証では、学習 e ポータルのアカウントについては「表示名」という自由度の高いデータ項目を用いて名寄せを試みており、個人を識別できる確実な情報を用いて照合したわけではないため、名寄せの結果については必ず正しいとは言えない。あくまで本実証の名寄せ結果は『候補』という扱いである。他人の ID と紐づいている可能性もあることを考えると、この結果をそのまま最終結果として ID を紐付けて運用するには大きな問題がある。

このため、名寄せ結果が誤っていないか必ず教育委員会、学校経由で本人に確認を取る必要がある。学習 e ポータル上のアカウントに対して個人を識別することのできる明確な情報が提供されない限り、本作業は必ず行わなければならない作業であると言える。

4 まとめ

本実証の結果から、今後、同様に発生するシステム間の名寄せ作業に活用できるようベストプラクティスを以下のとおりまとめる。

4.1 今後の活用に向けて

4.1.1 名寄せに必要な条件

名寄せに必要な条件は、まとめると以下の通りである。

I. 名寄せに必要なデータ項目があること

少なくとも以下のいずれかが設定されたデータ項目が学習 e ポータルと校務支援システムの双方に必要であるといえる。

①出席番号

②姓名(漢字 or ふりがな or ローマ字)

また正当性を高めるために ①と②のデータと合わせて、③生年月日などを組み合わせて確認し、可能な限り重複することなく一意に特定できる状況にできるとベストである。

II. 名寄せに必要なデータ項目の形式が統一され、必須登録となっていること

効率的かつ正確に名寄せを行うためには、必要なデータ項目を保持するデータ形式がシステム内でルール化されているとベストである。(必要なデータ項目があったとしてもその値が自由に設定できるとツール等を用いて判定する際により多くのパターンで比較を行わなければならない、エラーが発生しやすくなるため。)

4.1.2 学習 e ポータルのデータ管理と名寄せの手法について

本実証においては、学習 e ポータル側に出席番号、姓名、生年月日といった、個人を識別できる確実な情報が保持されていないこと、また「表示名」という自由度の高いデータ項目を用いて名寄せを行うしかなかったことから、汎用的な名寄せツールが使えず独自のツールでより多くのパターンで比較をしなければならず多くの工数がかかった。

また、比較処理が複雑になればなるほど名寄せツールの作成と品質維持が難しくなり、名寄せの失敗も発生しやすくなる。それによって人力での確認や名寄せにも多くの時間を要することとなり、ヒューマンエラーの発生確立が高まることで最終的な名寄せ結果の信頼性が低下する恐れもある。

学習 e ポータル事業者ごとにデータ保持の形式は異なるため、連携先の事業者によっては名寄せの難易度も変わってくると思われるが、いずれの学習 e ポータルにおいても、個人を特定できる情報を持たないというポリシーのもと運用されているケースが多く、確実な名寄せ結果を得る難しさは常にあると考えられる。これらを回避するためには、「4.1.1 名寄せに必要な条件」にも記載した通り、学習 e ポータル側で個人を識別できる確実な情報を統一されたフォーマットで保持することが不可欠だといえる。

今回のようにシステム間で連携を行う機会は今後も増えることと考えられるが、その際に初回限りとはいえ名寄せ作業の度に、多くの工数をかけ信頼性の低い結果しか得られないことを考えると、学習 e ポータル内にも出席番号や姓名、生年月日といった確実に個人を一意に識別できる何かしらのデータを保持しておくことが最善の策であると考えられる。一方で学習 e ポータル内に個人識別用のデータを整備し登録する作業についても工数がかかると思われるが、ある程度統一された形式で個人識別用のデータを保持することで、汎用的なツールの活用が可能となり、名寄せの工数をおさえるとともに信頼性の高い名寄せの結果を得ることができると考えられる。

以上