

Emigrant Industrial Savings Bank Test Book Accounts Entries, Linked Data Representation, 1850-54

Creators

Nicholas Wolf (ORCID: 0000-0001-5512-6151)

Wesley Chioh (ORCID: 0000-0002-9742-7144)

Marion Casey

Patricia Feighery

Kevin Rich

Acknowledgements

Initial work on refining and identifying extended family members in the original transcription records was completed with support from a New York University Office of the Provost University Research Challenge Fund.

Subsequent support to transform the data into a linked-data format was completed with assistance from a grant from the New York University Center for Faculty Advancement Curricular Development Challenge Fund.

License

The files presented here are provided under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>). Users are free to download, transform, copy, and redistribute the data contained in the files, provided that the original data are cited, the subsequent derived data is shared under the same provisions, and the data are not distributed for primarily commercial purposes.

Citation

Nicholas Wolf, Wesley Chioh, Marion Casey, Patricia Feighery, and Kevin Rich, "Emigrant Industrial Savings Bank Test Book Accounts Entries, Linked Data Representation, 1850-54," New York University Faculty Digital Archive, 2020, <http://hdl.handle.net/2451/61521>.

1. Introduction

1.1 Overview

This dataset contains a specific representation of a subset of the savings-accounts test books of the nineteenth-century Emigrant Industrial Savings Bank (EISB), modeled using a linked-data format. The EISB was established in New York City in 1850 by the Irish Emigrant Society. An institution marked by its community of depositors representing a broad section of the city's immigrant population of all classes, particularly its rapidly growing Irish-born population, the opening of over 25,000 accounts between 1850 and 1860 generated a variety of paper records ranging from the test books in question here, which recorded individual information about depositors, to real-estate records and bank-balance ledgers.

The test-book depositor records provide researchers with an insight into the immigrant origins of New York City in the 1850s. Beyond a microeconomic account of everyday banking, the data provide examples of the occupation held by residents of the city in the mid-nineteenth century, supply attestation to various addresses in the city via geocodable residential locations, and, most significantly, through the information provided in an unstructured column in the test books labeled "remarks," a free-text descriptive overview of the transatlantic home locations of family members who had migrated to across the globe. Among its potential analytics uses, addresses can be georeferenced to recover the neighborhood locations of immigrants, their home townlands, parishes, and cities in Ireland and other countries (most notably, Germany). Demographic, occupational, and familial features of depositors are also available, as are the means by which they arrived in New York City by ship. Most of these depositors were recent immigrants hailing from Ireland, the United Kingdom, and the constituent states of the then German Confederation.

The original EISB records were obtained by the New York Public Library after being rescued from destruction by NYU historian Marion Casey. These records were microfilmed for

preservation and, more recently, transferred to digital images files for inclusion in the NYPL Digital Library. This includes the test files in question here, available at <https://digitalcollections.nypl.org/items/33091bd0-fcba-0132-7758-58d385a7bbd0>. The EISB records have also been the focus of two projects aiming to convert its structured data into data files for computing. The NYPL's *Emigrant City* project (launched ca. 2015 and formerly at <http://emigrantcity.nypl.org/>), sought to crowd-source the transcription of the bank's mortgage and bond records. The second, an NEH-funded effort led by researchers Anbinder, Ó Gráda, and Wegge to transcribe the test-account books and link their person-level microdata to a variety of nineteenth-century records (census and ship records in particular) resulted in a CSV-structured, wide-format, tabular dataset published on Harvard's Dataverse instance (see <https://doi.org/10.7910/DVN/DDG8SW> for the latest version).

The basis for this present, linked-data version of the dataset is a subset consisting of accounts 1-4008 of a larger transcription effort spearheaded by Kevin Rich and published in print form as the multivolume *Irish Immigrants of the Emigrant Industrial Savings Banks* (2001-2015). Rich's transcriptions presented the account entries in a tabular form in Excel files, with a single row for each account holder. Crucially, adjunct data for the account holders' relatives, present in the "remarks" column, had not been parsed into separate persons records in Rich's original version; extended family information beyond siblings, spouses, parents, and children that are available in this field are also omitted in the Anbinder et al. data.

1.2 Linked-Data Model Justification

Previous transcription efforts, which constrain themselves to a strictly tabular presentation of the accounts, provide one path toward modeling the data presented in the EISB accounts. However, two unexplored avenues are presented here as helping to analyze this data in new ways. The first involves the unstructured "remarks" field referenced above, the second the use of linked data built on the Resource Description Framework (RDF; subject-property-object format) standard in order to better accommodate the diverse supporting data sources that can be mixed with the EISB records to enrich our knowledge of the individuals recorded in the bank books.

Like a miscellaneous notes field, the "remarks" column in the original test books attempted to encapsulate heterogeneous information that could help identify next-of-kin, credit references, and other relevant facts collected about depositors. This included, most notably, nativity information, but also the names of siblings, children, spouses, in-laws, grandparents, and cohabitating persons of depositors. Small details like languages spoken, locations of relatives in

a global context, and even personal health circumstances were included. However, the information was entered more or less in free-text style, without organizing atomic information into separate columns. Prolific use of abbreviations, plus the general difficulty in making out handwriting, can make for considerable challenge in using the field. Here, for example, is one such “remarks” column entry:

Native of Cormack, 7mi from Loughrea, Co Galway Ire - Arr NY May 07 per the ship Howard from L'Pool - Parents Dead, Fa Owen- Mo Mary Quill, 2 Sisters Marg in NY- Bridget in Ire -Is Single

This entry tells us that the depositor is from Cormack, probably a townland, located 7 miles from the town of Loughrea, County of Galway, Ireland; that the depositor arrived in New York on May 7 on the ship Howard from Liverpool; that the depositor’s parents are dead, that the depositor’s father was named Owen, mother Mary Quill, and that the depositor has two sisters, one in New York named Margaret and one in Ireland named Bridget who is unmarried. The information in this field is thus rich, but difficult to recover in an organized, machine-readable fashion.

Work conducted in 2016 to 2018 was able to separate out the information in the “remarks” field for the first 4,008 accounts using human-based (but machine-assisted) examination of each record. This resulted in additional records (in total, 29,099 persons recorded in the data) for relatives and household members to accompany main account holders:

Persons Referenced in the Test Books:	29,099
Total Number of Accounts:	4,008
Account Opening Years:	1850 - 1854

It is this refined dataset that provides the basis for the linked data triples provided in this iteration. Two advantages can be offered for providing this data in a linked-data version. First, as with all linked-data objectives, it separates out the various facts encoded in the data for easier ability to link these discrete facts with related assertions about these individuals provided in other data sources such as census, newspaper, civil, and religious records. Separating out the birthplace of individuals, for instance, allows new datasets providing information about those locations to be very quickly incorporated into the EISB data while avoiding any ambiguity about the places represented.

Second, this format helps to further the process of establishing firmer identities for the place names and familial relationships listed in the data. The EISB records, even in their original form,

were prone to clerical spelling errors or misheard names recorded in the account books. The spelling of some place names were incorrectly Anglicized, or else reference an entity (such as pre-unified German states) that no longer exist. This project aims to provide a more solid identity for these locations by georeferencing these mentioned places wherever possible.

2. Data Structure

2.1 Subjects Identifiers

Each individual mentioned in the EISB test account books has been assigned a unique identifier, an “esbid,” regardless of whether that individual is an account holder or not. This identifier is derived from the account number (which exists in the original and is widely used by other datasets derived from the same original sources such as the Anbinder et al. study) combined with a two-digit sequential appended number. Those final two digits have been incremented for each person mentioned (usually family members) in conjunction with the main account holder. Each account number is padded with the leading number 1 plus sufficient zeros so that all esbid numbers consist of 8 digits. Thus:

Example 1

Account Number on File:	1
Total Number of Individuals Named in the Account Entry:	11
Range of esbid values:	10000100 - 10000110

Example 2

Account Number on File:	3182
Total Number of Account-Holders:	1
Range of esbid values:	10318200

Example 3

Account Number on File:	235
Total Number of Account-Holders:	2
Range of esbid values:	10023500 - 10023501

These unique identifiers, in the format esbid:XXXXXXXX (for example, <esbid:1000100>) form the basis of most subjects in the RDF triple stores provided here, and often objects as well. Normally, RDF would require the subjects of this dataset to be a fully formed URI, but in the absence of a deployment of these data on the web at this time, this “URI-like” format has been used. Using this identifier, the associated account number of any individual can be immediately retrieved by removing the leading 1 and 0s, as well as the final two digits.

In cases where an individual is not directly mentioned in the account entries, but is inferred from the description of other family members for whom the undescribed person is essential to conveying the nature of the familial relationship between the persons present, an “inferred” person was added to the familial group and a special appended flag, the string “_inf,” was added to the esbid number (for example, <esbid:10178201_inf>). See below under family relations for more information.

Note that while these identifiers uniquely refer to a person’s appearance in the context of a given account, those persons may be mentioned multiple times in conjunction with multiple accounts across the span of the depositor books. For instance, if an individual opened one bank account, then in a later year opened another, that person will appear more than once in this dataset under two different identifiers.

2.2 URI Authorities Used for Properties and Objects

The following linked-data authorities have been utilized to provide URI for property and object values:

Wikidata (<https://www.wikidata.org>)

Logainm.ie (<https://logainm.ie>)

Geonames (<https://sws.geonames.org>)

All properties and objects describing personal features other than spatial ones rely on Wikidata. The other two authorities have been used for locational attributes in the following priority order:

1. Given that most of the account holders are Irish, an attempt was first made to find an URI in the place-name authority Logainm.ie, the official data provider for Irish locations (Logainm.ie is itself a linked-data project).
2. If the spatial location being referenced was not an Irish one, Geonames was used. All country-level URIs, including that of Ireland itself, were drawn from Geonames.
3. If the location no longer exists in either Logainm or Geonames, a historical location from Wikidata was supplied.

Another authority was used for object properties of occupations, namely the classification system for occupational categories developed by scholar Robert Ernst in Appendix VII of his *Immigrant Life in New York City, 1825-1863* (1949; Syracuse University Press, 1994). Although not able to provide URI values for RDF objects, the Ernst categories do provide a restricted taxonomy that helps group diverse occupational terms into a more compact description. Thus, a separate file lists Ernst occupational categories as string-value objects of all persons listed with occupations in the dataset.

2.3 Properties

2.3.1 Family Relations

Relevant Files:

- family_relations.nt
- extended_family_relations.nt
- extended_family_relations_crosswalk.nt

Count:

- 38,181 triples (family_relations.nt)
- 932 triples (extended_family_relations.nt)
- 30 triples (extended_family_relations_crosswalk.nt)

Information about familial relationships are found mostly in the “remarks” column of the original account entries and are oriented as statements of relationships to the main account holders. Often, these can be very extended, with entire extended family members, including sibling’s families and multiple generations, described. In an effort to make compact data files, any relationship between the main account holder and another individual that could be inferred from that account holder’s parent-child or spouse-spouse relationship to the main account holder were not encoded in a separate (and redundant) triple statement. Thus, sibling relationships, for example, are not directly described but can be reconstructed by linking two objects (children) whose subject (a parent) are the same.

2.3.1.1 Primary Relationships (parent-child and spouses)

Primary relationships are listed in the family_relations.nt file, and can be summed as follows:

1. Anyone described as a “Child,” “Daughter,” or “Son” of another person in the EISB accounts were placed in the object position of a triple, with the main account-holder person as the subject. The property is Wikidata P40, “child,” <https://www.wikidata.org/wiki/Property:P40>, so that a triple statement <parent-subject> <P40:property> <child-object> is used.
2. Anyone coded as “Father”, “Mother”, or “Parent” were placed in the subject position of a triple, with the main account-holder person as the object, and related by the Wikidata P40 property.
3. Siblings, i.e. those recorded as “Brother,” “Sister,” or “Sibling” are not directly described, but can be reconstructed by bringing together all child objects sharing the same parent-subject.
4. The spousal relationship relies on the Wikidata property P26, “spouse,” <https://www.wikidata.org/wiki/Property:P26>. No gendered order of preference has been deployed here in cases where both spouses were account holders; the first named account holder is asserted as the subject, the second as the object. In cases where a spouse is not named as an account holder, that spouse serves as the object.
5. Parents-in-law (mothers and fathers) can also be reconstructed here; they have been listed as the subject/parent of a person in the object position who is in turn also the object of a spouse triple; the subject of the spouse triple is the child-in-law of original parent-in-law.

In the event that no parent has been described in the account entry, but that parent is needed to clarify other relationships such as siblings or in-laws, then a parent record has been minted as an inferred person. The esbid of these inferred individuals will have an “_inf” suffix.

Note that there were cases of first spouses (i.e. missing or deceased spouses) recorded in the test books. These have been given triples as well, so that an individual may be described as being married to more than one person.

2.3.1.2 Extended Family Relationships

Wikidata does not currently have properties for every kind of familial relationship. Instead, it has a general property for a family relationship, which then receives a clarifying property type attached to that relationship; there are also some object properties that would identify a family

member (for example, there is a “sister-in-law” URI, <https://www.wikidata.org/wiki/Q3238556>). Thus, to reconstruct family relationships not captured by the base triples explained above, use a combination of the `extended_family_relations.nt` and the `extended_family_relations_crosswalk.nt` files. The first makes assertion of a “relationship” (`<rel_id:XXX>`) between two individuals (usually the account holder and some relative), and gives a relationship id type. From this base, the `extended_family_relations_crosswalk` file must be used to recover the exact nature of that relationship. For example:

Example 1

```
<esbid:10324800> <rel_id:103> <esbid:10324811> .  
<rel_id:103> <https://www.wikidata.org/wiki/Property:P1039> <https://www.wikidata.org/wiki/Q2914212> .
```

The first triple asserts that account holder Robert Gribbin (#10324800) is related to an unnamed person (#10324811), a man described in the remarks field as a brother of Robert’s wife who lives in Ireland. The second triple explains that relationship #103 is of the brother-in-law type.

Example 2

```
<esbid:10288700> <rel_id:107> <esbid:10288703> .  
<rel_id:107> <https://www.wikidata.org/wiki/Property:P1039> <https://www.wikidata.org/wiki/Q15312935> .
```

The first triple asserts that account holder Ellen Toomey (#10288700) is related to a man with the first name Larce, who lives in New York. The remarks field explains that he is a half-brother. Thus, relationship #107 is that of a half-brother.

2.3.2 Occupations

Relevant Files:

- `occupations.nt`
- `occupations_ernst.nt`
- `occupations_ernst_group.nt`

Counts:

- 3,838 triples (`occupations.nt`)
- 3,640 triples (`occupations_ernst.nt`)
- 3,611 triples (`occupations_ernst_group.nt`)

Occupations for some (but not all) account holders were given a separate column in the original account books. Here, those occupations have been given as strings in the object position of the triples contained in the occupations.nt file. Occasionally occupations were listed for other family members. When this happened, it is included here.

Those occupations have been matched whenever possible to a general occupational category according to the Ernst classification (occupations_ernst_group.nt), as well as to a detailed category (occupations_ernst.nt). Those categories, also as string values, provide the objects of the triples stored in occupations_ernst.nt and occupations_ernst_group.nt.

Examples:

Occupation:

<esbid:10172300> <<https://www.wikidata.org/wiki/Property:P106>> "Glazier" .

Ernst Detail:

<esbid:10172300> <<https://www.wikidata.org/wiki/Property:P106>> "Glazier" .

Ernst General:

<esbid:10172300> <<https://www.wikidata.org/wiki/Property:P106>> "Building Trades Workers" .

These triples assert that Laovin Lane (a Polish immigrant, ID number 10172300), was listed in the account books as a glazier. His detailed Ernst occupational category is “glazier”; his general Ernst occupational grouping is as a part of the “building trades workers” category.

2.3.3 Gender

Relevant Files:

- gender.nt

Counts:

- 29,933 triples

Gender has been inferred from the personal name and familial relationships of each individual. Wikidata URIs for gender categories have been implemented here.

Note that not all individuals have been listed with a gender triple. In some cases, especially with young children, a person was described simply as an “infant” with no gender listed.

There are more records here (29,933) than there are indexed persons (29,099) because inferred

persons, such as in-laws (see above) required to establish family memberships, are listed here.

Examples:

<esbid:10363708> <<https://www.wikidata.org/wiki/Property:P21>> <<https://www.wikidata.org/wiki/Q6581072>> .

<esbid:10005702_inf> <<https://www.wikidata.org/wiki/Property:P21>> <<https://www.wikidata.org/wiki/Q6581072>> .

In the first triple, it is asserted that Mary (#10363708), whose last name is unknown and who was a sister of the account holder, is a female.

In the second triple, it is stated that a female who belongs to the account 57 grouping has been inferred. We learn from the family relationship file that this is a minted record meant to represent the account holder's mother. No further information about the account holder's mother has been given in the bank records.

2.3.4 Personal Names

Relevant Files:

- first_names.nt
- family_names.nt

Counts:

- 27,417 triples (first_names.nt)
- 10,481 triples (last_names.nt)

First and family names were given for the main account holders; further name information, often first names only, were given for family members in the "remarks" column. Most significantly, maiden names for women were often provided. Here, where a maiden name was provided, this is the last name asserted in triples associated with a female (married names can be inferred via the information associated with a spouse record).

It can be expected that not all individuals in this data will have a name; family names in particular were often omitted, making retrieving such names especially tricky for women. Young children (who are sometimes just described as "child") too are underdescribed in terms of names.

Example:

<esbid:10281101> <<https://www.wikidata.org/wiki/Property:P734>> "McNamara" .

This triple refers to Thomas McNamara, a father of the main account holder for account #2811.

2.3.5 Addresses

Relevant Files:

- location_address.nt

Counts:

- 3,971 triples

The original account books recorded the person's location in a "residence" column. These often consisted of specific street addresses, but in some cases granularity could vary (e.g. "Troy, NY"). Locations and addresses have been presented here as string values, using the Wikidata P6375 ("street address") property. Addresses are more or less presented as they are in the original test books, with minor cleaning. For the most part, the test books presented addresses uniformly as street number followed by street name. However, in some cases this order in the original books was reversed.

Example:

<esbid:10027300> <<https://www.wikidata.org/wiki/Property:P6375>> "382 Grand St" .

Account holder Fenton McEvoy resided at 382 Grant St.

2.3.6 Nativity

Relevant Files:

- birth_counties.nt
- birth_countries.nt
- birth_parishes.nt
- birth_states.nt
- birth_townlands.nt
- birth_places.nt

Counts:

- 2,953 triples (birth_counties.nt)
- 7,432 triples (birth_countries.nt)
- 375 triples (birth_parishes.nt)
- 255 triples (birth_states.nt)
- 30 triples (birth_townlands.nt)
- 2,644 (birth_places.nt)

Birth locations for individuals were given in a variety of granularities; sometimes country only has been provided, in other cases not only an Irish townland but even a general location within the townland. Efforts have been made to capture all levels within this locational hierarchy using triples referencing Logainm.ie, Geonames, or Wikidata URIs.

If a location was Irish, the triple it will be georeferenced to the Logainm.ie database. The remaining places are first georeferenced using the GeoNames geographical database, and WikiData for historical places that no longer exist today.

If a birthplace was definitively labeled as a townland, and a townland corresponding to that name (within the correct county) could be found in Logainm.ie, it has been provided here in the birth_townlands.nt file. Otherwise, if an Irish place has been mentioned but not labeled a townland, it has been provided in the birth_places.nt file and pegged to the proper URI.

The birth_counties.nt file includes Irish, U.S., and British counties. The birth_states.nt file is used to encode U.S. and Canadian states.

Every effort has been made to obtain URIs for birthplaces (which in turn will speed georeferencing, since coordinates will be available at the place URI). The exception are Irish parishes, and in particular Roman Catholic parish names. Because the Irish civil parish naming system follows the Church of Ireland's parish names, the Roman Catholic parishes, where provided in the test books, have been listed here as string values if they deviate from the administrative parish.

Not all individuals will have birth location triples included for the following reasons:

1. No birthplace was mentioned in the test accounts.
2. The place name given does not match any of the known place names.
3. The place name given might refer to a different place with another name today, a fact that can be guessed at by looking at the descriptors in the "remarks" column such as distance from a landmark. Logainm.ie does not give historical names, so these Irish place name that do not have a URI have not been given.

- The place name given is non-unique, and without any additional information, it is impossible to accurately pinpoint the place in question.

Finally, there is one separate category of birthplace encoded here. Several individuals were listed as “Born at Sea.” These individuals have been listed in the birth_counties.nt files as having the place of birth object of <<https://www.wikidata.org/wiki/Q46998262>>, “Birth at Sea.”

Examples:

```
<esbid:10302600> <https://www.wikidata.org/wiki/Property:P19> <http://data.logainm.ie/place/100022> .
<esbid:10139501> <https://www.wikidata.org/wiki/Property:P19> <https://sws.geonames.org/2963597/about.rdf> .
<esbid:10207800> <https://www.wikidata.org/wiki/Property:P19> <http://data.logainm.ie/place/1605> .
<esbid:10017901> <https://www.wikidata.org/wiki/Property:P19> <https://sws.geonames.org/5128638/about.rdf> .
<esbid:10330300> <https://www.wikidata.org/wiki/Property:P19> <http://data.logainm.ie/place/39144> .
<esbid:10086700> <https://www.wikidata.org/wiki/Property:P19> <http://data.logainm.ie/place/41827> .
```

Person #10302600 was born in the County of Limerick.

Person #10139501 was born in Ireland.

Person #10207800 was born in the civil parish of Ballingarry.

Person #10017901 was born in the U.S. state of New York.

Person #10330300 was born in the Irish townland of Fordstown.

Person #10086700 was born in the Irish townland of Loughan.

To assist with interpreting the birthplace information encoded in these triples, crosswalks of URIs to the raw strings have been provided in files counties_uri.csv, country_uri.csv, parish_uri.csv, places_uri.csv, states_uri.csv, and townland_uri.csv.

3. File Listing

Intermediate Data Filename	Description
counties_uri.csv	Maps names of counties to URIs from geographical databases
country_uri.csv	Maps names of countries to URIs from geographical databases
parish_uri.csv	Maps names of parishes to URIs from geographical databases
places_uri.csv	Maps names of places to URIs from geographical databases
states_uri.csv	Maps names of states to URIs from geographical databases
townland_uri.csv	Maps names of townlands to URIs from geographical databases

Primary Data Filename	Description
extended_family_relations.nt	RDF of extended family members
extended_family_relations_crosswalk.nt	Maps names of countries to URIs from geographical databases
family_relations.nt	RDF of immediate family members
sex_gender.nt	RDF of each account holder's recorded sex
occupations_ernst_group.nt	RDF of each account holder's recorded occupation grouped by Ernst Occupational groups
occupations_ernst.nt	RDF of each account holder's recorded occupation classified by Ernst
occupations.nt	RDF of each account holder's recorded occupation
location_address.nt	RDF of each account holder's recorded address
birthcountries.nt	RDF of each account holder's recorded birth country
birthcounties.nt	RDF of each account holder's recorded birth county
birthstates.nt	RDF of each account holder's recorded birth state (US or Canada)
birthparishes.nt	RDF of each account holder's recorded birth parish (Ireland or UK)
birthplaces.nt	RDF of each account holder's recorded birth place
birthtownlands.nt	RDF of each account holder's recorded birth townland (Ireland only)
last_names.nt	RDF of each account holder's recorded last name
first_names.nt	RDF of each account holder's recorded first name