

Linking records of early aeronautical experimenters across data sets

By Peter B. Meyer, U.S. Bureau of Labor Statistics¹
For session on Innovative Techniques in Record Linkage
SSHA conference, Chicago, November 17, 2016
Preliminary and incomplete; please do not cite.

Abstract. This paper describes an experimental wiki database containing records about aeronautics and aviation globally up to 1916, during the period when the idea of flying machines transitioned from a dream and a hobby into a new science and a startup industry. Some data series extend over this 1800-1916 period.

Over 14,000 patent records have information about the applicant, and over 30,000 bibliographic records of publications have some identification of the author. Smaller datasets cover early aeronautical clubs, aviation firms, exhibitions and conferences, and letters between aeronautical enthusiasts during this time period. At least 1000 individuals appear in these data sets or are otherwise thought to be significant to the invention and early development of airplanes. Our research links these records from multiple sources to construct individual and organizational histories and will eventually yield networks of co-authorships and partnerships. One challenge is that we do not have a unified definitive source of purely biographical information, such as birth date or full name. We put these diverse kinds of records on a public wiki so that specialists who are not part of the initial project can improve the underlying data and the matching.

This paper describes an experimental database system which maintains links between data records that cover an extended historical period. In the period, category systems are changing and the source records have problems of ambiguity and uncertainty – that is, the historian cannot always ascertain whether two persons or documents are the same, or which one is referred to in a particular primary source or secondary historical work. The design of the database and its interface is intended to enable careful memory of historical facts and conclusions, and to support the use of human decisions about each record or document, while enabling the creation of statistical measures from the data as it evolves and grows.

The database is available as a wiki at <http://aero.referata.com>. A *wiki* is a web site that shows data records and allows users to edit the records directly from their web browser. Such systems are auditable in the sense that each past edit to a record is remembered, and it is straightforward to check who made an edit and when, and to compare the versions before and after the change. In a wiki, the records are easily hyperlinked to one another, and properties of these links are also usable data, for example, a link might be from a record about a patent to one of several category systems to classify what the patent is about.

¹ Views and findings in this work do not represent the Bureau of Labor Statistics. Many persons and institutions have contributed to the data shown here, including Intan Hamdan-Livramento and Julio Raffo of WIPO.

The subject matter of this database is the long period of development of aeronautics into aviation from the early 19th century to 1916. Ballooning was widely recognized by 1800, but balloons could not be piloted well. In 1809 a journal article included the first known published design for a fixed-wing aircraft. The idea developed adherents, and by 1870 ballooning clubs included members attentive to this “aerial navigation” idea – *aeronautics*. A variety of designs were modeled and explored. It wasn’t until 1903 that a fixed-wing aircraft could demonstrably carry a person and be controlled in flight. An industry of airplane-makers developed starting in 1908 across the industrial countries. Throughout this long period and up to 1916, the database has tables of data on patents, on ballooning and aeronautical clubs, on the new startup firms, and on exhibitions of aeronautical models. It will also have data on publications about aeronautics, ballooning, and aviation. Each item can be classified in a number of ways, and its network relations to other items recorded and developed with hyperlinks. This helps historical conclusions be grounded straightforwardly to evidence.

Patents and patent categories

I have collected data on patents related to early aeronautics and aviation (Meyer 2013, 2015; Meyer *et al.* 2016). The sources have been the sites of the US Patent and Trademark Office, the European Patent Office, Google patents, and dozens of historical sources which identify a patent as relevant to flight. Now these are combined with several thousand records from the World Intellectual Property Organization, the UN agency that manages patent rights and treaties. Most patents were filed in France, Britain, Germany, or the U.S.

An example patent document is shown in Figure 1. The invention shown is an important glider designed by Otto Lilienthal of Berlin, with wings roughly in the shape of those of a soaring bird. The patent filing illustrates certain elements of most 19th century patents: a number in the national government’s ordering scheme, a title, the patentee’s name and location or citizenship, the dates it was applied for (“filed”) and granted, and a diagram. It is only two pages long, shorter than most patents of the time and much shorter than patent filings now. Its title is “Flugapparat,” meaning “flight apparatus.” Patent officials classified it by its technology, but the classification in this case is “Sport” -- no aeronautics category existed.



PATENTSCHRIFT

— № 84417 —

KLASSE 77: SPORT.

OTTO LILIENTHAL IN BERLIN.
Flugapparat.

Zusatz zum Patente Nr 77916 vom 3. September 1893.
Patentiert im Deutschen Reiche vom 29. Mai 1895 ab.
Längste Dauer: 1. September 1908.

Bei dem unter Nr. 77916 geschützten Flugapparat hat sich der Liebelstand gezeigt, daß, wenn der Apparat die Luft unter sehr spitzem Winkel durchschneidet, die Vorderkante infolge der gewölbten Flächenform Druck von oben erhalten kann. Dadurch wird ein stabiles Durchgehen der Luft gefördert, und der Apparat aus seiner Flugrichtung gedrängt.

Um dieses zu vermeiden, wird die vordere Flächenpartie drehbar gemacht, das die Vorderkante drehbar sich nach unten richten kann. Das in Fig. 1 schraffierte Flächenstück kann sich um die Achse *a b* nach unten, etwa bis in die Lage *e d* (Fig. 2) herabsenken, durch einen Luftdruck von unten aber wieder bis in die Lage *c e* erheben. Durch federnde Organe *ff* hat das schraffierte Flächenstück das Bestreben, die geneigte Lage *e d* einzunehmen, und zwar ist der normale, auf diese bewegliche Fläche einfallende Luftdruck gerade ausreichend, um die Federn *ff* so weit zu spannen, daß das vordere Flächenstück in die

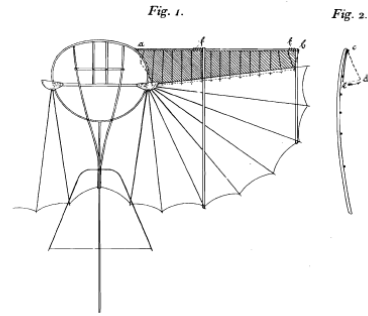
gehobene Lage *c e* gelangt und dadurch ein Theil der ganzen geschlossenen Flügelfläche wird. Hierdurch ergibt sich die Wirkungsweise insofern, als bei einer Luftdruckverminderung unter der schraffierten Fläche *c e* die federnden Organe die Fläche selbst nach unten drücken, wodurch der verminderte Luftdruck sich wieder ergänzt und aufrichtend auf den ganzen Apparat wirkt, bis die zu einem stabilen Fluge des Apparates erforderliche Lage wieder erreicht ist.

PATENT-ANSPRUCH:
Eine Ausführungsform des durch Patent Nr. 77916 geschützten Flugapparates, bei welcher der vordere Theil der Flügelfläche um die Vorderkante (*a b*) nach unten drehbar ist und durch federnde Organe *ff* nach unten gedrückt wird, so daß er sich beim Nachlassen des von unten wirkenden Luftdruckes nach unten dreht und dadurch ein den Apparat aufrichtendes Moment erzeugt.

Hierzu 1 Blatt Zeichnungen.



OTTO LILIENTHAL IN BERLIN.
Flugapparat.



Zu der Patentschrift
№ 84417.

PHOTOG. DRUCK DES REICHSDRUCKERS.

Figure 1. Otto Lilienthal's 1893 patent

Identifying which patents are relevant to early aeronautics is a general problem in this research, for several reasons. First, it is not easy to get general lists of all patents of the time; the U.S. is the easiest country for this as it has all patents back to 1837 online. Second, the technological classifications of the time do not clearly identify aeronautics, and the classifications differ by country. (In the case of Lilienthal's patent, the title includes the term for flight, which is a good clue, and the patentee was famous for his accomplishments in this area so a museum and several books have listed this patent.)

Third, there is the underlying uncertainty about which patents had implications for aeronautics at a time when it was not known that the soaring-bird model of an aircraft would be more successful than balloons, rockets, helicopters, or flapping-wing (*ornithopter*) designs. Specific historical research can evaluate this but a simple search or download cannot. There is deep uncertainty associated with potential new technologies.

Patents after about 1910 appear in the EPO (European Patent Office) database called PATSTAT and can be searched through their web site Espacenet.com. Almost all of these patents have been classified by technology. For identifying earlier patents, contemporaneous evidence is available (e.g., Brewer and Alexander, 1893). Figure 2 shows a list of patents published in a journal which were identified by contemporaries as relevant to aeronautics.

Figure 2. Aeronautics-related French patents listed in a 1901 issue of the journal *L'Aérophile*

| LISTE DES BREVETS | |
|--|--|
| RELATIFS A L'AÉRONAUTIQUE ET AUX SCIENCES QUI S'Y RATTACHENT | |
| DEMANDÉS EN FRANCE DU 10 AOÛT 1901 AU 19 SEPTEMBRE 1901 (1) | |
| 313,652. | — 20 août 1901. — de Dion : Perfectionnements aux ballons dirigeables. |
| 313,665. | — 21 août 1901. — de Dion : Perfectionnements à la construction des ballons dirigeables et à leurs mécanismes de propulsion. |
| 313,675. | — 21 août 1901. — Sébillot : Perfectionnements dans la navigation aérienne. |
| 313,676. | — 21 août 1901. — Duguet : Aérostat dirigeable. |
| 313,682. | — 24 août 1901. — Diennin : Nouveau système de ballon dirigeable. |
| 313,689. | — 22 août 1901. — Hébert : Planophile le « Georges Hébert », nouvel appareil d'aviation. |
| 313,758. | — 25 août 1901. — Lemoine : Perfectionnements aux aéroplanes. |
| 313,796. | — 26 août 1901. — Delaurier : La navigation aérienne pyrotechnique. |
| 313,957. | — 2 septembre 1901. — Turr : Nouveau système de ballon. |
| 313,962. | — 3 septembre 1901. — Papiér : Perfectionnements dans l'aérostation. |
| 313,995. | — 4 septembre 1901. — Porak : Machine à voler. |
| 314,095. | — 7 septembre 1901. — Vroland : Système de ballon dirigeable. |
| 314,146. | — 10 septembre 1901. — Riedinger : Hélice aérostatique à poids formant volant. |
| 314,207. | — 12 septembre 1901. — Magnié : Nouveau ballon dirigeable. |
| 314,219. | — 13 septembre 1901. — Guillaume : Perfectionnements apportés aux dispositifs employés dans la navigation aérienne. |
| 314,368. | — 19 septembre 1901. — Piatti dal Pozzo : Aérostat dirigeable. |

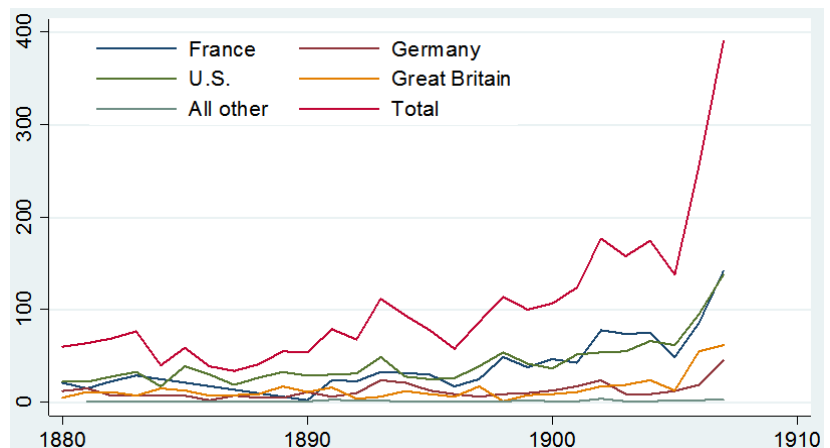
(1) Communication de MM. Marillier et Bohélet, Office International pour l'obtention de brevets d'invention en France et à l'étranger, 32, boulevard Bonne-Nouvelle, Paris.

PARIS. — IMPRIMERIE CHATELAIN, 7, RUE BEUCLER. Le Directeur général : GEORGES BESANÇON.

Another set of ambiguities is that some patents report single original inventions, but these cannot always be readily distinguished in the sources from two secondary kinds of patent: (a) an “addition” to an earlier patent, that is, a clarification or new claim, which itself is usually called a patent, and (b) the foreign filing of a patent already granted in a first country. These latter patents may look like new patents in the data but generally do not represent the same inventive significance. Nowadays a set of such closely related patents is called a patent family. A record linkage goal for this wiki system is to enable visibility and discovery and easy linking of these secondary kinds of patents back to the first one and vice versa.

Until 1907 patent counts overall and in aeronautics in particular grew exponentially by 5% a year or more in these countries as shown in Figure 3. Then there was a spike, with the appearance of the new industry, followed by a decline from about 1913 into World War I, as the startup wave ended and secrecy became more important. A fuller analysis calls for classifying these into a common technology category system, distinguishing at least basic categories of balloons, helicopters, flapping-wing ornithopters, fixed-wing airplanes, and cross-cutting categories.

Figure 3. Patents per year related to aeronautics and aviation before the airplane industry began, using the relevance criteria in the text.



Personal names are often incomplete in old patent records. One might see just the surname, as in Figure 2. Sometimes this too is ambiguous as in the case of an inventor listed in PATSTAT as “DON SIMONI.” It turned out that was entirely a surname; his full name was François André don Simoni. Worse, sometimes only one inventor is listed in patent data when there were three or four inventors. Furthermore there is routinely ambiguity in data sources between the inventors and the applicant, or the assignee, who might be someone else or a company with some ownership interest. For a technological history, one wants to distinguish these, as the original records usually do. Google patent has clear PDFs of all the relevant U.S. patents but their character-recognition scan of the patent docs is imperfect. Record systems of early patents also tend to have misspellings. For all these reasons the researcher may have to clean the data, not only download it and use it. Even in modern U.S. patent data, there is not always certainty about whether one patentee is the same as another, because the patentee is required to give a name but no further identification, and, as I understand it from US PTO chief economist Alan Marco, this problem is worse in data before 1975. New algorithms are being put into place to probabilistically identify inventors across patents (Monath, McCallum, *et al.* 2015). A parallel effort imputed location data of U.S. patentees probabilistically (Petralia, Balland, and Rigby, 2016). That data, called HistPat, is incorporated into the data on the wiki.

Thus data on patents which should in principle be clearly countable is in fact fraught with uncertainty, ambiguity, and is not always available. This project confronts those issues of clarity directly and naively. Such records as we can obtain are downloaded or transcribed and typed into a system which never loses track. When we resolve a name, such as don Simoni’s, it stays resolved in the system along with whatever evidence was used to resolve it. If information is probabilistic, such as the information in HistPat which classifies the locations of U.S. inventors, we can include it and resolve it by hand over time.

An example patent record on the wiki is in Figure 4. It includes a structured list of fields, followed by “wikitext” which can describe the patent less formally and link to outside sources including, when available, the patent documents themselves at Espacenet, Google patents, or the national patent offices. The structured table can include links to inventors by name, and to technological categories. These links have the “semantic” property that they can be used by other pages; notably, this one which refers to inventor Octave Chanute can be listed on the page about Chanute. The link to Augustus Herring is red because a page for him does not yet exist.

| Patent GB-1897-15221 | |
|-------------------------------|---|
| Year filed | 1897 |
| Year granted | 1897 |
| Office | GB |
| Patent number | 15221 |
| Inventors | Octave Chanute, Augustus M. Herring |
| Inventor country | US |
| Applicant person | |
| Applicant firm | |
| Applicant type | |
| Applicant is inventor? | Yes |
| Original title | Improvements in or relating to means and appliances for effecting Aerial Navigation |
| English title | Improvements to means and appliances for effecting Aerial Navigation |
| Tech fields | aeroplane |
| Filing date | May 31, 1897 |
| Full specification filed data | |
| Application number | |
| Grant date | June 25, 1897 |
| Granted? | Yes |
| Publication date | |
| Supplementary to patent | |
| Related to aircraft? | |
| Serial number | 1 |
| Patent agent | Thomas Moy |
| Assigned to | |
| National tech categories | |
| IPCs | |
| CPCs | |
| Family year | 1897 |
| First filing? | No |
| Number of citations | |
| Cited by | |
| Application ID | |
| INPADOC family ID | |
| Number of text pages | |
| Number of diagram pages | |
| Number of figures | |
| Number of claims | |

It is a triplane with engine, as well as a biplane, with two or three stacked wings; two propellers, one in front and one in rear, rotating in opposite directions, according to Simone Short and Steve Spicer at Spioenweb, which says the patent application filed 31 May 1897 through Thomas Moy as the agent for O. Chanute and A.M. Herring, British Patent No. 15,221 (C.P.L.)^[1] Neilson (1910), p24, 73 also refers to it.
Espacenet does not have this patent document online as of November 2016.

References

- The main source for this record is An Extended Bibliography on Octave Chanute [#] at Spioenweb

Category: Patents

Figure 4: Example patent record on the wiki

Organizations – clubs and firms

Balloon clubs existed by 1800, and over the course of the next century more appeared and a growing number of their members were interested in “aerial navigation” – that is, aeronautics. By 1870 some distinctive organizations focused not on ballooning but on aeronautics and aviation. There does not seem to be a single encompassing global database or history of these organizations. Historical narratives sometimes refer to a member organization but not clearly, or use a translated name so it is ambiguous which original organization was referred to. Their dates of starting, ending, reincorporation, and reorganization are not central to the usual narratives, which means basic information on these organizations is not clear. Often an organization had multiple branches or chapters, or a cluster of distinct organizations had a common federation. So far we have records of several hundreds of these, most of which were founded as local chapters in the later years of the sample period.

Starting in 1908 a wave of new companies manufactured airplanes. For many of these startups we can identify the founders and a rough starting date but the details of incorporation are not clear yet. We have names or other information on over 500 such companies around the world, including some duplication still. A shakeout probably took place starting around 1911, but the evidence is not clear. It is often unclear when new companies disappeared or merged so it is difficult to construct a population history of the new industry. A key designer may be reported as being at one firm, then at another, with no clear statement of whether the first one still exists. This requires historical investigation, linkage of different sources, and disambiguation. Records of the best-known information can be kept on the wiki. I have chosen to use English translations of organization names in the wiki page titles, and to list the names under which the organization was known on that page.

Exhibitions

Balloon exhibitions started before 1800, and there were related aeronautics sections and model demonstrations. These were important institutions in that they encouraged people who went on to make valuable contributions to aeronautical engineering. Some exhibitions were famous, such as the World Fairs which had aeronautical demonstrations. Yet there does not seem to be a global database of them, and their details are frequently unclear and ambiguous. We have information – at least an event year and name – on 70 of them.

Publications

Bibliographies of publications about aeronautics are available (Brockett, 1910, 1921). These have been partly digitized and when converted to databases will be uploaded. Then notes on each historic article can be kept, and links to the authors can be set up. Often the bibliography entry does not list the author’s full name so there will be a lot of reconciliation to do, and some residual uncertainty will remain, but the publication can be remembered online in any case. The bibliographies have about 13,000 entries up through mid-1909, and an estimated 25,000 more through 1916.

Inventors, authors, and letters between them

With combined information on publications, patents, and the other materials mentioned it will be possible to construct better systematic records on those individuals who did not become famous but did participate in the long slow development of the airplane. We can show better and clearer records of patent counts, publication counts, membership in clubs, attendance at exhibitions, and references in historical works. Such statistics give ways to summarize the efforts or contribution of some of these experimenters and scientists, and the numbers can be expanded, improved, and sharpened with information from the wiki. It is of general interest how often these individuals participated in group activities, and how often they published outside their country borders. It seems so far to have been a lot, meaning the technology advanced internationally, not separately across languages and borders.

This approach usually does not lead to specific rules or probabilistic decisions, but rather to disambiguation by human investigation, case by case. Presumably related patents and publications around the same year with similar but ambiguously-named authors are by the same person and a deeper investigation is conducted. Outside sources are welcome in the search and can be cited on the wiki pages. When there is uncertainty about the decision, it stays in the records and can be reevaluated later. The service of computer scientists here is not data science, but to give us careful record-keeping and a usable interface in which no one is a bottleneck holding the one master copy of the data. In the future the data can give us network measures of distance based on coauthorships or partnerships, but first the basic historical matching of individuals and organizations is needed.

Many letters and telegrams between early aeronautical experimenters are available because the experimenters were literate, separated geographically, and the search for a functional flying machine took decades. Substantial published volumes have the correspondence of the most important inventors – the Wright brothers, of Octave Chanute, and of Otto Lilienthal (McFarland, 1953; Schripps, 1985). Much more is in archives. Such correspondence identifies and clarifies many other people who made some contribution. The wiki has data on 600 letters so far, generally including a sender, recipient, and date. Often the text is available online.

It appears that at least 1000 individuals who appear in the wiki attempted to make some contribution in written form to early aeronautics and aviation. Until further record linkage and disambiguation is done the true numbers are not clear. The industry employed thousands of people during World War I.

Wiki systems for historical research

Historians wish to identify, disambiguate, and clarify inventors, authors, firms, patents, clubs, and so forth. An existing historical narrative may be entirely persuasive, but if the source evidence is transparent, findings and claims are more reproducible and alternative lines of thought can be evaluated more directly.

One thing the wiki can do is to combine information on individuals into a common database. This aero wiki has a special software extension called *Cargo* which enables wiki pages to conduct queries and get reports on the other wiki pages if they have certain structured information in them (Koren, 2016). Since the pages on patents, or on letters, are clearly marked, this means the page about an individual can have an automatically updated report from the patents or letters. Figure 5 shows part of the page on the most prolific early aeronautics writer, Octave Chanute.

Figure 5. Automatically generated reports listing wiki pages that make certain semantically distinct references to the author and inventor Octave Chanute: patents, publications, and letters he wrote

Octave Chanute

Octave Chanute (b. 1832 in Paris) was one of the most important aeronautical inventors and networkers before the invention of the airplane.^{[1][2]}

References [edit]

- ↑ Octave Chanute on English Wikipedia
- ↑ Simine Short. 2014. *Locomotive to Aeromotive: Octave Chanute and the Transportation Revolution*

Patents by Octave Chanute

Patent GB-1897-15221, Patent GB-1898-13372, Patent GB-1898-13373, Patent US-1897-582718, Patent US-1906-834658

Publications by Octave Chanute

Chanute, 1893. The secret of soaring

Letters sent by Octave Chanute

Octave Chanute to G. Corelli 12-Oct-1897, Octave Chanute to G. Corelli 21-Apr-1898, Octave Chanute to G. Corelli 22-Nov-1898, Octave Chanute to G. Corelli 6-Jun-1898, Octave Chanute to John Montgomery e, Octave Chanute to John Montgomery g, Octave Chanute to John Montgomery h, Octave Chanute to John Montgomery i, Octave Chanute to John Montgomery j, Octave Chanute to John Montgomery l, Octave Chanute to John Montgomery m, Octave Chanute to John Montgomery n, Octave Chanute to Louis-Pierre Mouillard 1-Jul-1892, Octave Chanute to Louis-Pierre Mouillard 1-Oct-1892, Octave Chanute to Louis-Pierre Mouillard 10-Oct-1894, Octave Chanute to Louis-Pierre Mouillard 11-Feb-1892, Octave Chanute to Louis-Pierre Mouillard 12-Dec-1892, Octave Chanute to Louis-Pierre Mouillard 12-Jan-1897, Octave Chanute to Louis-Pierre Mouillard 12-Jul-1891, Octave Chanute to Louis-Pierre Mouillard 12-Sep-1892, Octave Chanute to Louis-Pierre Mouillard 13-Aug-1893, Octave Chanute to Louis-Pierre Mouillard 14-Jun-1893, Octave Chanute to Louis-Pierre Mouillard 15-Apr-1896, Octave Chanute to Louis-Pierre Mouillard 15-Aug-1894, Octave Chanute to Louis-Pierre Mouillard 15-Jun-1894, Octave Chanute

Reports on patents and letters are generated automatically by a one-line query statement in the underlying wikitext that will be translated into an SQL query for the database underlying the wiki and the result then shown in the browser when the user loads the page.

```
{{#cargo_query:table=Patents|fields=_pageName|where=Inventors HOLDS 'Octave Chanute'|format=ul}}
```

Thus the wiki enables systematic, detailed information on the individuals, and invites the reader to drill down into detail and identify network relationships. A historian might show a letter and say Chanute imitated, or followed the pattern of, an earlier aircraft design. This wiki can enable a more complete understanding. Evidence on the existing networks helps address network questions and what-if question such as: what if there were no Wright brothers? When would there have been an airplane?

Similarly the wiki can keep detailed information on patent technology categories. Figure 6 shows a sample page. The international classification system called CPC (Cooperative Patent Classification) categorizes aeronautical patents in detail but was invented long after the period under investigation. However a page about each patent category can identify its subcategories and supercategories and similar categories in other patent systems. These pages like other wiki pages can have footnotes and link to other pages and include pictures.

The screenshot shows a Wikipedia page for the patent category 'CPC B64C11/00'. At the top, there are navigation tabs: 'page', 'discussion', 'edit with form', 'edit', 'history', 'delete', 'move', and 'protect'. Below the title, there is a table with the following data:

| | |
|----------------------|--------------------------|
| Enclosing categories | CPC B64C |
| Subcategories | CPC B64C11/46 |
| Affiliated concepts | CPC F01D5/06, propellers |
| Notes and sources | |

Below the table, there is a paragraph of text: 'Propellers, e.g. of "ducted type"; features common to propellers and rotors for rotorcraft.^{[1][2]} The phrase "ducted propeller" means is a propeller whose nozzle does not rotate; another name is Kort nozzle.^[3]'

Underneath, it says 'This class covers:' followed by a bulleted list:

- Propeller hubs, blades and pitch-changing mechanisms^[2]
- Propeller vibration absorbing or balancing means; Arrangements of multiple propellers (e.g. coaxial propellers).^[2]
- Active or passive propeller measures for noise reduction (only such disclosures are attributed the symbol B64C 11/00).^[2]

There is a 'References' section with three entries:

- ↑ B64C11/00 elements ↗ listed at Espacenet
- ↑ 20 21 22 23 B64C11/00 ↗ listed at USPTO
- ↑ Ducted propeller ↗ on English Wikipedia

At the bottom, there is a 'Category' tag: 'Techtypes'.

Figure 6. An official patent category as described and footnoted by the wiki's editors

To edit this wiki is a bit tricky still and there are many details to correct. It is slower to edit than a spreadsheet. It has the huge advantage that the work can be decentralized and volunteer efforts can simply be accepted without careful management until a problem is observed. It will also offer downloads of the full data to any user, at some point, and hopefully volunteer specialists will wish to contribute information.

The wiki tracks all its edits which means that if an editor is making edits that are not valuable, this can be quickly discovered by looking at the history of the page itself, or the history of a user who made an edit. These are both easy to see and visible to anyone. Looking at the edits of others is a straightforward way to learn to use the system and make edits oneself.

A wiki system has the useful feature that an arbitrary number of “redirect” and “disambiguation” pages can be created. So if a user creates a wikilink to USPC or CPC, users clicking on these links will be directed automatically to those pages. If a user searches for “Paris exhibition” – an ambiguous term -- the system can offer a list of several exhibitions held in Paris in the relevant time period.

Conclusion

This paper characterizes a number of uncertainties, ambiguities, and unknowns in the data on early aeronautics and aviation of the period 1800-1916. The wiki system addresses these by maintaining links between records of different kinds, and with some semantic elements so that the links are not all of the same kind – they have predefined meanings. The wiki can straightforwardly link back to sources, either online or by bibliographic reference, and thus it supports specific historical research which can improve aggregate data sets. Each observation can be tracked in multiple category systems, e.g., the multiple patent technology classification systems, which have to be harmonized to draw conclusions about the frequency of balloon-related patents relative to fixed-wing ones over time.

In the wiki system, the data are open and available to query and download. Edits to the system are tracked which provides some provenance information on changes and helps ensure reliability. It also means that the system does not completely forget information it once had; past information is in the page’s history. Misspellings or duplicates can be discovered and corrected by simple searches. If a misspelling is corrected, the existence of the earlier misspelling can be noted in the wiki text. Duplicate records in the system are common but can be combined by hand.

Other historical researchers can benefit from this kind of platform. In research on classes of individuals with many qualitative differences between them, such as early women investors as a class, or inventors in the Industrial Revolution, has this character. A narrative of each person and links to related structured data can be kept together, and statistics drawn from recent historical discovery.

Bibliography

- Brewer, Griffith, and Patrick Y. Alexander. 1893. *Aëronautics*. London. Reprinted in 1965 by Boekhandel en Antiquariaat, B. M. Israel N.V. Amsterdam.
- Brockett, Paul. 1910. *Bibliography of Aeronautics*. Smithsonian Institution.
- Brockett, Paul. 1921. *Bibliography of Aeronautics, 1909-1916*. Smithsonian Institution.
- Koren, Yaron. Cargo extension to MediaWiki. <https://www.mediawiki.org/wiki/Extension:Cargo>
- McFarland, Marvin W., ed. 1953. *The Papers of Wilbur and Orville Wright*. Vols. 1–2. New York: McGraw-Hill.
- Meyer, Peter B. "The Airplane as an Open Source Invention." *Revue Economique* 64:1, (January 2013), pp. 115-132.
- Meyer, Peter B. "A Catapult of Riches: the invention of the airplane and its industry." 2015. Chapter 8 of *Institutions, Innovation, and Industrialization: Essays in Economic History and Development*, edited by Greif, Kiesling, and Nye. Princeton U Press.
- Meyer, Peter B.; Intan Hamdan-Livramento, Julio Raffo. 2016. Aeronautical patents and aviation history from 1880-1916. Presentation at ASSA 2016, SGE session.
- Monath, Nicholas; Andrew McCallum with Michael Wick, Jack Sullivan and Ari Kobren. Discriminative Hierarchical Coreference for Inventor Disambiguation. Presentation to PatentsView workshop, USPTO, 24 Sept 2015.
<http://www.patentsview.org/data/presentations/UMassInventorDisambiguation.pdf>
- Petralia, Sergio; Pierre-Alexandre Balland; David L. Rigby. 2016. Unveiling the geography of historical patents in the United States from 1836 to 1975. *Nature*, 30 August 2016.
<http://www.nature.com/articles/sdata201674>
- Schwipps, Werner. 1985. *Collected Correspondence of Otto Lilienthal*.