https://doi.org/10.14195/2182-7974_35_1_1

The student records of the University of Coimbra (1537-1919): an Open Data Science approach

O ficheiro dos alunos da Universidade de Coimbra (1537-1919): uma abordagem de ciência de dados aberta

JOAQUIM RAMOS DE CARVALHO
Coordinator of the International Portuguese Training Centre
Macao Polytechnic University
joaquimcarvalho@mpu.edu.mo
ORCID: http://orcid.org/0000-0003-2325-7098

Artigo entregue em: 11 de maio de 2022
Artigo aprovado em: 13 de junho 2022

ABSTRACT

The University of Coimbra keeps to this day the academic records of students since 1537. In the years 1940-50 a card file of student information was created, known as the "Ficheiro de Alunos". The catalogue contains records from 1537, when the university was relocated from Lisbon to Coimbra, up to 1908. The amount of information varied over time, containing name, first and last date on record, place of origin, school, years of enrolment, degrees obtained, and results of exams or other proofs of proficiency. Many records also contain notes such as titles (for instance indicating nobility), religious order, and college of residence. In the years 2013-15, the contents of the card files were input into an archival management system, giving the old records a new digital life. Currently, around 105,000 records are available

online, and reachable through search engines. This paper addresses two limitations of the current online catalogue: first, the academic information in the paper cards was transcribed as a single text field preventing the usage of structured queries and any type of non-trivial data analyses; secondly, the opportunity of the University Archive to improve the catalogue through the cooperation of its users lacks a collaborative model that can scale. This article contributes solutions to both issues: we present algorithms to extract information from the records and produce representations in line with current data science paradigms, allowing a wide range of interesting analysis of the data; we also demonstrate how tools and cooperation models developed in the open-source community can provide an environment for collaborative efforts ranging from the notification of simple errors to the addition of semantic web representations for linked data, harnessing the knowledge dispersed by many researchers working on this unique repository of data. All source code and data analysis produced for this paper are available in a public repository at https://github.com/joaquimrcarvalho/fauc1537-1919.

KEYWORDS: University of Coimbra; Students; Database; Open Science; Data Science.

RESUMO

A Universidade de Coimbra reserva até hoje os registos académicos dos seus alunos desde 1537. Nos anos 1940-50 foi criado um catálogo de fichas de papel, conhecido por "Ficheiro de Alunos". O catálogo contém registos desde 1537, quando a universidade foi transferida de Lisboa para Coimbra, até 1908. A quantidade de informação variou ao longo do tempo, contendo nome, primeira e última data de registo, local de origem, faculdade, anos de matrícula, graus obtidos, e resultados de exames ou outras provas de proficiência. Muitos registos também contêm notas como títulos (por exemplo, indicando nobreza), ordem religiosa e colégio de residência. Nos anos 2013-15, o conteúdo do catálogo foi inserido num sistema de gestão de arquivos, dando aos antigos registos uma nova vida digital. Atualmente, cerca de 105.000 registos estão disponíveis online, indexados nos motores de busca. Este artigo aborda duas limitações do atual catálogo online: primeiro, as informações acadêmicas nas fichas de papel foram transcritas em um único campo de texto, impedindo consultas estruturadas e qualquer tipo de análise de dados não triviais; em segundo lugar, as oportunidades de melhorar o catálogo através de contributos dos seus utentes carece de um

modelo colaborativo sustentável. Este artigo contribui com soluções para ambas as questões: apresentamos algoritmos para extrair informações dos registros e produzir representações alinhadas com os paradigmas atuais da ciência de dados, permitindo uma ampla gama de análises interessantes dos dados; também demonstramos como ferramentas e modelos de cooperação desenvolvidos na comunidade de código aberto podem fornecer um ambiente para esforços colaborativos que vão desde a notificação de erros simples até aplicações de web semântica e dados ligados, aproveitando o conhecimento disperso por muitos pesquisadores que trabalham neste repositório de dados único. Todo o código-fonte e análise de dados produzidos para este artigo estão disponíveis em um repositório público em https://github.com/joaquimrcarvalho/fauc1537-1919.

PALAVRAS-CHAVE: Universidade de Coimbra; Estudantes; Base de Dados; Ciência Aberta; Ciência de Dados.

Introduction

In the wider context of research in the history of universities, the University of Coimbra constitutes a special case for three reasons: the preserved series of academic records on individual students, from 1537 to the present, which allow for studies over long periods; the fact that Coimbra was the only general-purpose higher education institution in the Portuguese empire until the early 20th century, meaning that those records refer to people who had important roles in many parts of the globe where the Portuguese were present; finally, its unique role in Portuguese history and early European globalization produced many flows of people from outside Portugal that came to Coimbra for a variety of reasons and went on to be important agents of European presence in the world, the Jesuits being the most relevant example.

A further specificity of the University of Coimbra is the existence of a centralised historical archive that doubles as the state archive for the Coimbra administrative area. The dual nature of the *Arquivo da University de Coimbra* (AUC) has contributed to the development of a stable institutional environment. The director of AUC is appointed by the Rector of the University of Coimbra; notable professors with a strong interest in the history of the university and information science have held the post and promoted cataloguing and description of the collection and the development of search tools.

The continued commitment of directors and staff produced, over a long period of time, the "Ficheiro de Alunos" (FA), a set of tens of thousands of paper cards, with information on students from 1537 to 1919, and the parallel description of the collections from which the information was gathered. The FA has become an invaluable tool for locating information about specific people, providing in most cases enough information for identification purposes and allowing the retrieval of the original sources with academic information.

Today the content of the card files is available on the web. The cards were transcribed into an archive management software using what could be called a semi-structured notation. It is now easy to locate a specific person of interest, with search engines handling queries for names and "Universidade de Coimbra" by presenting results directly from the FA (see Figure 1).

Figure 1 – Example of a name search linking directly to "Ficheiro dos alunos".

However, due to the textual representation of the academic information in the digital version of the catalogue, it is not possible to do structured queries on the information, such as "students entering the university in 1789", "students from Brazil before 1822", "students present at the same time as Eça de Queiroz", and so on.

The work presented here is intended as a contribution to the development of the tools that can foster the development of studies about the history of the University of Coimbra.

Two questions will be addressed: first, is it possible to produce a version of the FA that will allow queries such as those exemplified above, based on data models compatible with data science tools and interoperability standards such as semantic web and linked data; second, how to design an environment that will ease the cooperation of researchers and citizens in the improvement of the FA, using a model inspired by open-source projects.

The software developed for this work is released as open-source on GitHub, including many listings exemplifying the contents of the FA[1].

History of the student records file

The catalogue of the students of the University of Coimbra, also known as "Ficheiro dos Alunos", was produced between 1940 and 1950, under the direction of Mário Brandão, an historian who carried out extensive research on the history of the University of Coimbra and who was, at the time, director of the University archives — *Arquivo da Universidade de Coimbra [AUC]*.

An overview of the production of the file records and their content is available as part of the online catalogue (*Índice de alunos da Universidade de Coimbra*, s.d.). A more complete guide for researchers can also be obtained online (Bandeira, s.d.). Figure 2 shows an example of a paper record.

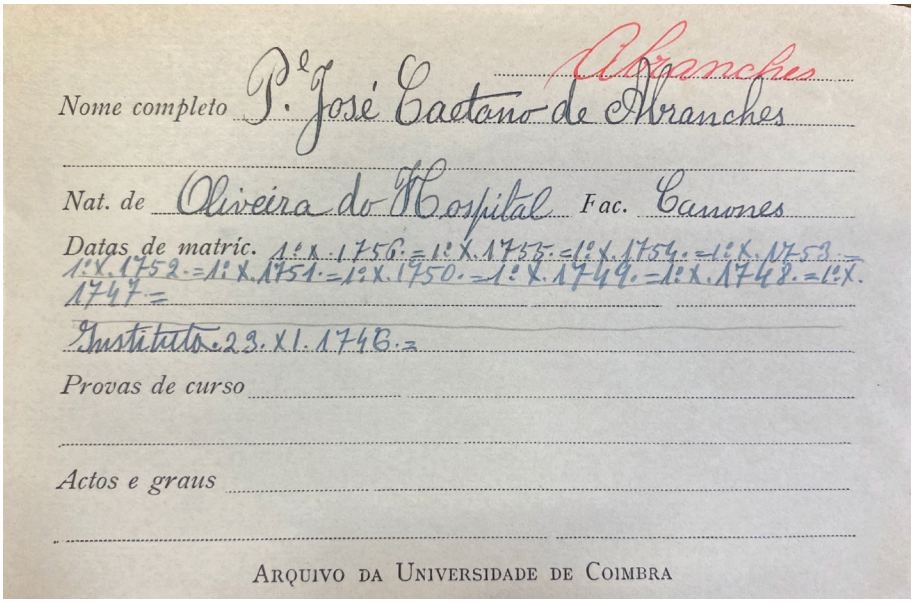
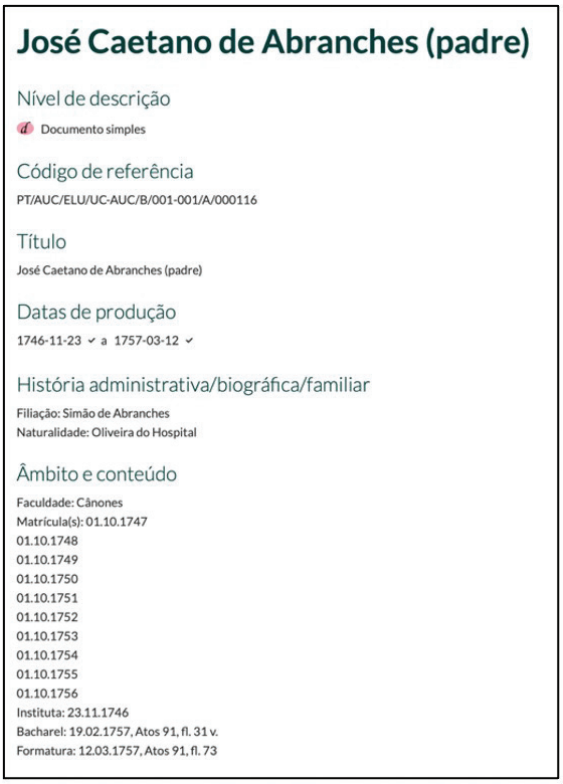


Figure 2 – Example of a file card. Note that pre-printed cards are used which predefine specific items: name, place of birth (nat. de), faculty (Fac.), dates of matriculation, proofs of attendance, acts and degrees.

[1] https://github.com/joaquimrcarvalho/fauc1537-1919. The repository includes the source code of what here is referred to as the "algorithm" and the Jupyter notebooks used to generate the data supporting this paper, as well as nominal listings of selected variables and statistics.

Between 2013 and 2015, the content of the paper cards was transcribed to an archival management software, which publishes the online version of the original paper cards[2]. From the moment it was converted into digital form the FA became a live catalogue, receiving updates, corrections and additions, diverging progressively from the paper cards.



José Caetano de Abranches (padre)

Nível de descrição

Documento simples

Código de referência

PT/AUC/ELU/UC-AUC/B/001-001/A/000116

Título

José Caetano de Abranches (padre)

Datas de produção

1746-11-23 a 1757-03-12

História administrativa/biográfica/familiar

Filiação: Simão de Abranches
Naturalidade: Oliveira do Hospital

Âmbito e conteúdo

Faculdade: Cânones
Matrícula(s): 01.10.1747
01.10.1748
01.10.1749
01.10.1750
01.10.1751
01.10.1752
01.10.1753
01.10.1754
01.10.1755
01.10.1756
Instituta: 23.11.1746
Bacharel: 19.02.1757, Atos 91, fl. 31 v.
Formatura: 12.03.1757, Atos 91, fl. 73

Figure 3 – Example of the online version of paper card of figure 1 (https://pesquisa.auc.uc.pt/details?id=140436): note that the information on degrees, absent in the original card, has been added to the digital record.

The software can export the content of the series in machine-readable format. A "CSV" (*comma separated values*) export file was used, dated February 2020 (with an update in 2022), provided by AUC[3].

2 (Archeevo – Archival management software, s.d.).

3 The *Archeevo* system can also export in EAD-XML format, which is more convenient for data exchange with other archival management software. See 'EAD: Encoded Archival Description (EAD Official Site, Library of Congress)' at https://www.loc.gov/ead/ [accessed 21 March 2022].

Methodology

A set of algorithms was created to process the exported information. The most important operations involve parsing semi-structured text with academic information. The Python library *pyparse* was used with a specially developed grammar to identify fields and values, handle variations in date formats and process names with annotations between parenthesis and specific constructs such as the word "vide" to denote cross-references. After a first pass to process meta-data and organise the text information into fields, a second pass applies a set of extractor functions that infer personal and academic information through the application of rules and normalizations. After the inference and normalization of information, a mapping phase generates records for a SQL database, which then supports the more complex forms of analyses presented here. Finally, the results are compared with published research for validation.

The ISAD-G records

The exported information follows the ISAD(G) model (International Council on Archives Conseil International des Archives, 2000). The student records correspond to a "Sub-series" (ISAD level: SSR). The Sub-series is divided into "units of installation" corresponding to the letters A to Z (level: UI). Each unit of installation contains the students whose last name starts with the corresponding letter, in alphabetical order of the last name.

Each student record in the export file has a six-digit identification number, specific to the *Archeevo* software, and a complete identification code, according to ISAD(G) specifications.

In the example of Figure 3, the complete identification code is PT/AUC/ELU/UC-AUC/B/001-001/A/000116. The internal identification code is present in the URL of the online record: https://pesquisa.auc.uc.pt/details?id=140436[4].

The export file contains a first row with the SSR record, one row for each letter of the alphabet (no entry for "Y"), and a row for each student. Figure N shows the content of an ISAD record as extracted from the *Archeevo* CSV export (empty fields and control information removed).

[4] To refer the reader to specific records in the text, we will use the internal six-digit code, which allows access to the online record at https://pesquisa.auc.uc.pt/details?id=nnnnnn, by replacing "nnnnnn" with a specific code. In the open-source repository associated with this article, we provide a mapping table between the full ISAD-G identification code and the 6-digit code used at the time of writing.

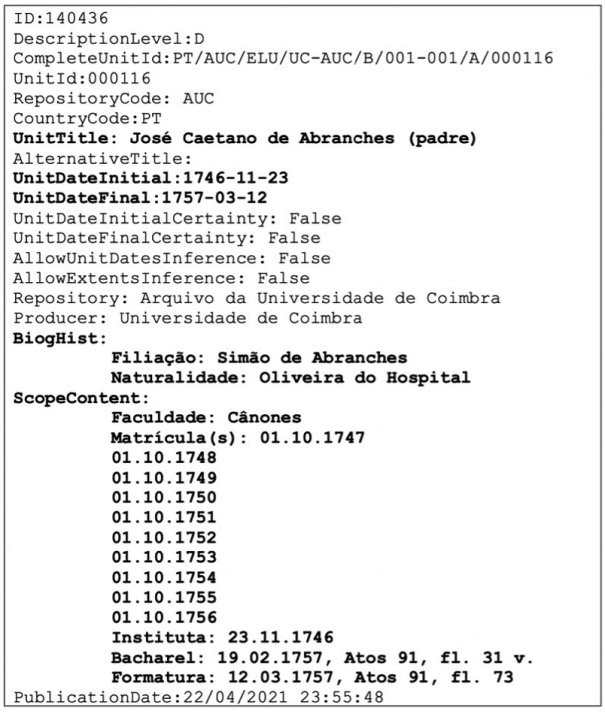


Figure 4 – Content of an exported record with ISAD-G fields. In bold face are the fields that contain useful information for historical analytical purposes.

Relevant fields

For analytical purposes the relevant fields are:

- UnitTitle = Name of the student, which can contain an annotation between parenthesis and a "vide" note, which is a cross-reference to another record.
- UnitDateInitial = first date on the record[5].
- UnitDateFinal = last date on the record.
- BiogHist = contains the place of birth ("naturalidade") and the name of the father (in a few cases also the name of the mother).

5 The fields *UnitDateInitial* and *UnitDateFinal* are not specified in ISAD(G) directly but allow support for structured data representations which are part of the EAD specification, see *Encoded Archival Description Tag Library Version EAD3 1.1.1* (Technical Subcommittee for Encoded Archival Standards of the Society of American Archivists, 219AD), p. 444 (tag: unitdatestructured).

- ScopeContent = contains the academic record with the information on faculty, matriculation dates, proofs of attendance and "acts and degrees" of the paper cards. This is a semi-structured text field, with the type of information, or field, suffixed with ":" followed by the content, but with many variations in the structure, as explained in the text below.

Extracting information from "ScopeContent" is the main challenge of processing the student records. Names of students and "BioHist" are comparatively simpler to handle.

Number of records

The file exported in February 2022 contained 105,298 student records.

There is a further reference in the file to 49,916 fathers and 93 mothers[6]. Some of the fathers are former students themselves, but there is currently no systematic way to detect those cases.

The relation of the number of records in the FA to the number of real students that studied at the University of Coimbra is affected by three factors: (1) the presence of cross-reference records, (2) the existence of duplicates and (3) possible sub-registration in the FA of students referred to in the base sources.

These issues affect, of course, all operations on the digital FA, putting limits in the type of data processing that can be done and the quality of the results obtained. They deserve careful analysis.

Cross-references and duplicates[7]

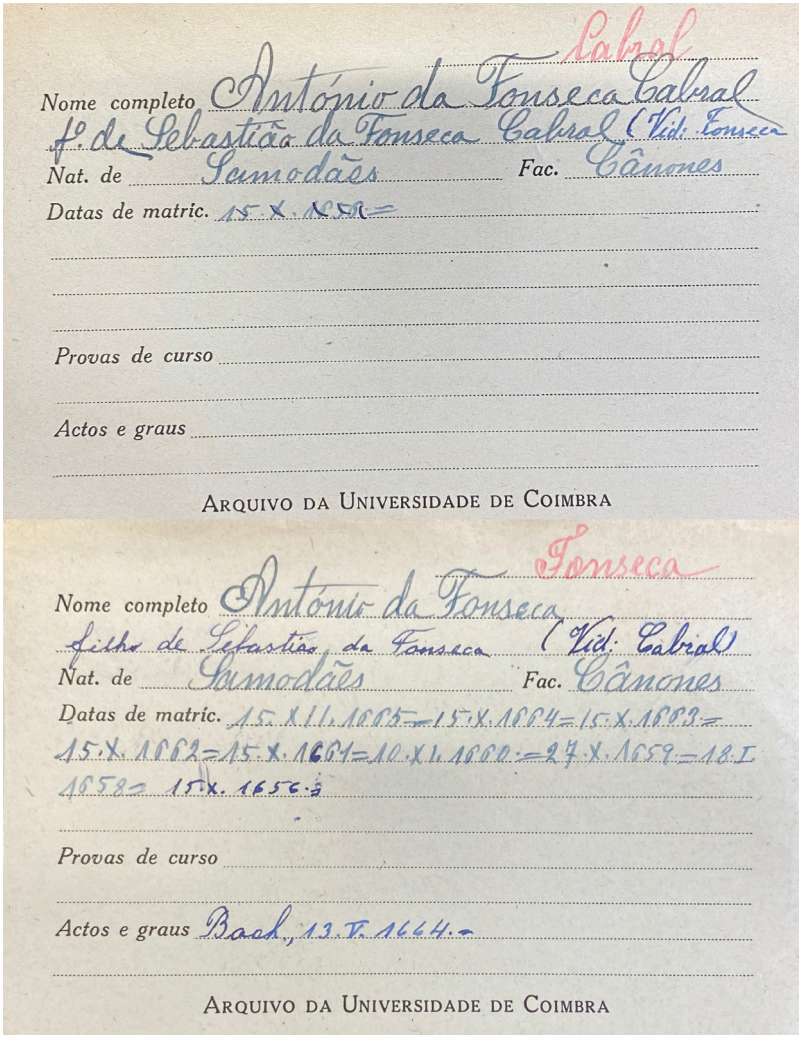
The digital version of the FA carries over a cross-reference mechanism from the card catalogue.

Figure 5 is rather revealing of the original scheme. It shows two of the paper cards linked by cross-references. Clearly, two moments occurred, distinguishable by different handwriting and ink colours on the cards.

6 Ana Maria Bandeira, responsible for the university records in AUC, clarified that the original books do not contain mothers' names. When, in the scope of specific research, a mother's name was established, it was added to the record. The FA is continuously evolving.

7 Data for this section was generated with https://github.com/joaquimrcarvalho/fauc1537-1919/blob/main/notebooks/015-remissivas.ipynb

Originally two cards were produced. At a later moment, they were found to pertain to the same student. The "vide" references were added compactly, since space in cards is limited. The father's name from the top card was copied to the bottom card. The single enrolment in the card on the top was crossed out over and added to the enrolments on the second card. In the digital FA two records were generated with names *"António da Fonseca Cabral,* ***vide*** *Fonseca"* and *"António da Fonseca,* ***vide*** *Cabral"*.


Cabral

Nome completo António da Fonseca Cabral

f.º de Sebastião da Fonseca Cabral (Vid. Fonseca

Nat. de Semodães Fac. Cânones

Datas de matric. 15 X 1658=

Provas de curso

Actos e graus

Arquivo da Universidade de Coimbra

Fonseca

Nome completo António da Fonseca

filho de Sebastião da Fonseca (Vid. Cabral)

Nat. de Semodães Fac. Cânones

Datas de matric. 15.XII.1665=15.X.1664=15.X.1663=

15.X.1662=15.X.1661=10.XI.1660=27.X.1659=18.I

1658=15.X.1656.

Provas de curso

Actos e graus Bach., 13.V.1664.

Arquivo da Universidade de Coimbra

Figure 5 – A cross-reference card pair in the original catalogue (now records 128053 and 139670). The "see" record is on top and the full record below. The record on top has the date crossed out with the same ink used in the "vide" expression. Also note that the "vide" expressions, although compact, allow the name of the other side to be inferred; since the rest of the information matches, the link is unambiguous.

These two types of records indicate that the original cross-reference scheme of the FA was designed to solve duplication issues and was bi-directional.

When the digital version was produced, the names with the associated "vide" expression were appended to the name in the "Unit Title" field of the ISAD-G record.

There are 8625 records with "vide" in the name, 8% of the total number of records. This number is significant enough to merit detailed inspection on how it can affect further data analysis.

Cross-reference expressions

The "name expression" after the word "vide", in most cases, is not a complete name, but rather a "hint" for a transformation of the "base" name. Here are some examples of names with "vide" expressions:

António da Fonseca Cabral, ***vide*** *Fonseca*
António da Fonseca, ***vide*** *Cabral*
Adriano Sisnando Brotero de Avelar Quintino, ***vide*** *Adriano Sisnando Brotero Quintino de Avelar*
Francisco António Campos, ***vide*** *de Novais Campos*

Types of cross-reference records

The records with "vide" in the name can be classified in two types, according to the type of extra information they contain (Table 1).

Type	Description	Number
Undated records	Almost empty records with a name with "vide" No dates (empty "UnitDateInitial" field) Other than the name: 93% place of birth 27% father's name 23% faculty	5704

Type	Description	Number
Dated records	Normal records with "vide" in the name Dates (valid "UnitDateInitial" field) Contain all types of information: 97% place of birth 53% father's name 99% faculty degrees, enrolment, and so on.	3062
TOTAL	Records with "vide" in the name (8% of total number of records)	8625

Table 1 –Types of cross-reference records in the digital FA.

The first type corresponds to what is normally considered a "see" reference: entries which contain very little or no information other than the name and the "vide" expression, and so direct the user "forward", to the full record under a different name. This is the case of the top record in Figure 5. And 93% of these "see" references contain name and place of birth.

The second type of records are full records. This is the case of the bottom record in Figure 5. In the digital FA they are only distinguishable from ordinary records because of the "vide" expression in the name. It is possible to pair them with matching records of the first type, as we will see below. Here the "vide" expression works as a "back reference" to alternate forms of the name, meaning "also known as" (aka), so that users of the card file who are locating references in the original sources would be informed of the alternative names under which the student appears in the books.

The scheme seems to be intended as bi-directional. If that were the case there should be a similar number of both types of records; however, we have a large imbalance (5563 vs 3062). For some reason, when creating a "see" reference, the "back/aka" was not always registered in the main record[8].

We also find several cases where the "vide" references links together full records. In these cases the meaning of the reference is "see also", and two records contain information about the same student, as we will see later.

Further insights on this can be obtained by trying to match the "vide" records in pairs, using the "vide" expression and place of birth as matching keys.

8 Some records contain multiple "vide" references, e.g.: 130281 *Nuno da Câmara (D.), vide Nuno Casimiro da Câmara e Nuno José da Câmara* which links with 130516 and 130517. At the time of writing 18 such cases were detected. The pattern of registering multiple cross-references varies, and the algorithm of extraction is not currently able to handle these cases.

Matching cross-references

A simple algorithm was used to find matching pairs of records with "vide" expressions. The goal was to test if the cross-reference scheme was indeed bi-directional and, more specifically, if the cross-reference scheme introduced duplicate records that could affect statistical and data analysis in general. No de-deduplication methods based on probabilistic record linkage were used, and although it can be shown that the FA contains unintentional duplicate records, that is a different endeavour[9]. Here we process records that were specifically marked as a cross-reference to another record by the producers of the catalogue.

The algorithm attempts to generate the "target" name that the "vide" expression implies. Then it will try to find pairs of records in which the "target" name of each one matches the name of the other, as in the example of Figure 5. The matching is restricted by place of birth, which seems to be the original strategy of the catalogue, considering that 93% of the "see" cards also contain place of birth.

The first step is to infer the target name from the "base" name and the "vide" expression.

The following four main transformation patterns were detected experimentally:

1. "Cut": *António da Fonseca Cabral, vide Fonseca*, result: *António da Fonseca.* The "vide" expression is a family name before the last; the target name is computed as the base name up to, and including, the "vide" expression.
2. "Add": *António da Fonseca, vide Cabral,* result: *António da Fonseca Cabral*. The "vide" expression is not present in the base name; the target name is the base name with the "vide" expression added at the end. In some cases, the linked record will have an extra particle before the "vide" expression, like "da", "de" and "do", which are common in Portuguese when aggregating multiple family names.
3. "Replace": *Adriano Sisnando Brotero de Avelar Quintino*, *vide Adriano Sisnando Brotero Quintino de Avelar*, result: *Adriano Sisnando Brotero Quintino de Avelar.* The "vide" expression is a full name. This hap-

[9] We did use "similarity" ratios when comparing place names, to allow for frequent variations in spelling.

pens when the transformation of family names cannot be expressed by "cut" and "add", so the producer of the card wrote the full target name after "vide" for clarity.

4. "Partial replace": *Francisco António Campos, vide de Novais Campos*, result: *Francisco António de Novais Campos.* The "base name" and the "vide" expression overlap at the end; the matched part in the "base name" is replaced by the "vide" expression.

"Cut" and "add" constitute 87% of the cases, with "replace" representing another 11%. Partial replacements are residual (26 cases).

Not all patterns in the FA are consistent. We find "vide" expressions that are meaningless, such as in record 205772: *Gaspar da Cunha Coutinho,* ***vide*** *Coutinho.* Partial replaces are also difficult to automate: 165045, *Isidoro da Cunha de Eça,* ***vide*** *dos Santos de Eça*, would generate *Isidoro da Cunha dos Santos de Eça*, but the actual target is *Isidoro dos Santos de Eça* in record 165046.

Applying these transformations, "target" names were generated for each of the 8625 records. Additionally, a version of the names without the Portuguese name particles ("de", "da", "do", "dos" and "e") was produced to facilitate comparisons[10].

Simple automated cross-reference resolution

The algorithm examines each record with "vide", generates the target name according to the transformations above and searches for a record with that name, the same place of birth and with a vide expression that matches the starting record. When all the records are processed, a check is made to detect ambiguous matches.

It was possible to match 3702 records in pairs by applying this set of transformations and checks (see Table 2). This represents 42% of all cross-reference records. More than half (1722/56%) of the full records with "vide" were matched with symmetric "see" records. Slightly more (1747) "see" records were matched with full records, but since there are many

10 The updated list of original names, "vide" expressions and generated target names is available at https://github.com/joaquimrcarvalho/ucalumni/blob/master/inferences/cross-references/vide_matched.csv. Listings in the repository are updated as the algorithms used here evolve and new analysis of the FA are performed.

more of those, the percentage was lower, 30.6% of the total. This was to be expected since there are, as explained before, more "see" records than "aka"[11].

Match result	Number of records	% matches	% of vide	% of vide type
Unambiguous matches	3702	97.9%	42.2%	
See to full record matches	1746	47.2%	19.9%	30.6%
Full record to see	1722	46.5%	21.4%	56.2%
Full record to full record	188	5.0%	2.1%	6.1%
See record to see record	10	0.02%	0.01%	0.18%
Records in ambiguous matches	116	0.3%	0.13%	3.0%

Table 2 – Results of cross-reference matching.

Additionally, the process also found 188 full records that matched other full records, 10 "see" to "see" matches and 116 "ambiguous" matches involving more than two records.[12] The 10 cases of match between see records are not relevant and are due to deficiencies in data input, or in the "vide" expression that triggers the algorithmic match.

Another question is the 188 (5%) full record matches. These, albeit few, testify that in certain cases the "vide" mechanism appears to have been used to link two separate full records that were produced separately and, at a later stage, found to pertain to the same person. Instead of creating a new card record consolidating the two, the cataloguer used "vide" expression to create a "see also" reference (see Figure 6). These records hint at a wider problem with duplicates.

11 A spreadsheet with the matching results is available at https://github.com/joaquimrcarvalho/fauc1537-1919/blob/main/inferences/cross-references/vide_matched.csv and another with the non-matched records https://github.com/joaquimrcarvalho/fauc1537-1919/blob/main/inferences/cross-references/vide_non_matched.csv.

12 Ambiguous matches occur when for a given "vide" more than one record was found, or when transitive matching occurs (record A matches B which matches C, etc.).

Simão Lourenço, vide Coelho

Nível de descrição
Documento simples

Código de referência
PT/AUC/ELU/UC-AUC/B/001-001/L/004757

Tipo de título
Atribuído

Título
Simão Lourenço, vide Coelho

Datas de produção
1648-10-31 ? a 1655-04-30 ?

Extensões
0 Álbum

História administrativa/biográfica/familiar
Filiação: Teodósio Lourenço
Naturalidade: Tancos

Âmbito e conteúdo
Faculdade: Leis
Matrícula(s): 1649.10.02
1650.10.12
1651.10.01
1652.10.31
Instituta: 1648.10.31
Examede Bacharel 1655.04.26
Formatura 1655.04.30

Simão Lourenço Coelho, vide Lourenço

Nível de descrição
Documento simples

Código de referência
PT/AUC/ELU/UC-AUC/B/001-001/C/009442

Tipo de título
Atribuído

Título
Simão Lourenço Coelho, vide Lourenço

Datas de produção
1650-10-12 ? a 1655-04-30 ?

História administrativa/biográfica/familiar
Filiação: Teodósio Lourenço
Naturalidade: Tancos

Âmbito e conteúdo
Faculdade: Leis
Matrícula(s): 12.10.1650
Instituta:
Bacharel em Leis 26.04.1655
Formatura 30.04.1655
Bacharel em Artes 14.03.1650

Figure 6 – Example of two full records of the same student, linked by "vide" expressions (194939, 250513). The record on the left starts in 1648 with the enrolment in the "Instituta" course, continues with enrolments in subsequent years and ends with the degree of "Formatura" in 1655. The record on the right starts with the enrolment in "Leis" in 1650, also in the other record, and contains a reference to "Bacharel em Artes", absent from the first record. Clearly, the two records pertain to the same student, and, if not merged, will introduce errors in detailed analysis.

Overall, the binary matches between "see" and full records are 93.7% of the matches made, which confirms that the original cross-reference scheme was indeed intended as a bidirectional link between non-preferred forms of the name and preferred forms of the name.

Nevertheless, with 42% of all the "vide" records matched the main question is why the remaining 58%, near 5000 records, failed to be matched.

It is known that there is an imbalance in the type of "vide", with many more "see" records than full records (5563 vs 3062). A manual search in the entire database starting with a sample of 20 unmatched "see" records, found 18 matches missed by the algorithm. In 14 out of 18, the reason for the miss was that the full record had no "vide" expression indicating the reverse link. Since the matching process was done only between "vide" records, the matching record was outside the search space. The other four cases of matches missed by the algorithm are due to variations or typing errors in the "vide" expressions or place of birth[13].

13 More detailed analysis and many examples can be found in the notebook referred in note 7.

Effect of cross-references on further analysis

So, to conclude the analysis of cross-references, it can be determined that a cross-scheme exists in the FA to register alternate forms of student names. It was the process used in the original card catalogue to resolve duplicates at an early stage. The scheme produced records that are "see" references, with little information other than name and place-of-birth, that "point" to the main record through a "vide" name expression. These "see" records are easy to detect because they have no values for the "Unit dates" fields. Over 5500 such records exist. In some cases, the full records to which those "see" records point also have a reverse reference, and it is not difficult to match automatically both types with a simple algorithm. A higher rate of matches would be obtained if all records in the matching process were included, and not just those with "vide" expressions, cleaning the many typing errors and variations in geographical names that exist in the dataset and using a more flexible algorithm, tolerant to variations in personal attributes. This was outside the scope of this research.

For the purpose of further exploration of the data it is sufficient to accept that by removing 5563 "see" cross-reference records from the data, we remove a considerable portion of the bias that the cross-reference scheme could introduce in a more fine-grained analysis of the FA. Those records would produce limited impact anyway, since they have no dates and only include, in significant numbers, information about the place of birth and to a lesser degree, father's name and "faculdade". Therefore, their impact in further analysis can be controlled.

More detrimental to the usage of FA is the existence of real duplicates: more than one record pertaining to the same student, as the example in Figure 6. 188 such records can be found in the cross-reference set, which means, potentially, 94 students can be counted twice. Although some of them seem to be "accidents" resulting from ambiguous "vide" expressions or "see" records to which dates were added by mistake, or not crossed out, and even completely identical records, we must assume that finding 94 duplicates in 8625 "vide" records hints that many more might exist in the full set of data. But how many more?

A pessimistic view on duplicates would be that 94 duplicates in 3062 (the number of full records examined for matches) is 3% and so, extrapolating, there could be around 3,000 duplicate records in the FA, after removing the "see" references. There maybe even more that this, considering that some of "ambiguous" matches most probably are due to "see" records matching duplicated full records.

An optimistic view could argue that the 188 pairs are the result of an ongoing effort to flag duplications with "vide" cross-references, and so there is an expectation that few remain undetected.

Moving forward it is clear that much can be done to detect duplicate records with the information in the FA and resolve the cross-references into matched records. At this point in time, I think that more sophisticated approaches to deduplication, using available tools in the data science toolset, should wait for further progress in data cleaning and insights provided by analysing the FA after extraction of structured information. Unintentional duplicates are more easily found in alphabetical listings of place-of-birth, father's name, degrees and other personal and academic attributes that can be produced. Also, it is possible that patterns of duplication emerge that allow some form of automated resolution.

To close this already long discussion on cross-references, the fundamental issue is the future of such schemes in a digital version of the FA, and if it should evolve into modern solutions for this type of problems, in the scope of ISAD(G) related standards.

Coverage[14]

The analysis of the cross-reference information in the FA allows for an informed, yet cautious, approach to the impact of duplicate records in data analysis.

But there is also the inverse question: To what extent can the FA be trusted to provide adequate coverage of the original sources?

To estimate the reliability of the FA it is necessary to refer to numbers or nominal lists that were obtained directly from the sources, without using the FA as mediation.

A first step is to chart the evolution of intake. This is relatively easy to do using the initial dates on the records. The "see" type cross-reference records do not show in this graph because, as explained previously, they have no date on record.

14 Data for this section was generated with https://github.com/joaquimrcarvalho/fauc1537-1919/blob/main/notebooks/020-students_overview.ipynb

Figure 7 shows the overall evolution of student intake from 1537 to 1917, the start and end dates on the FA. The question is how to assess the reliability of these numbers[15].

Figure 7 – Evolution of student intake. Yearly totals are obtained from "UnitDateInitial" field in the ISAD(G) record.

Chronological profile

A first attempt is to compare the chart above to similar charts derived from source data.

Fernando Taveira da Fonseca produced such a chart for the period 1577-1820[16]. We can compare it with a similarly configured plot with data derived from the FA.

15 For a thorough discussion of the issues related to the reconstruction of the student population of the University of Coimbra see chapter I of (Fonseca, 1995).

16 (Fonseca, 2007). According to the author: "The figures used here for 'freshmen' in Canon and Civil Law were obtained directly from archive records (Arquivo da Universidade de Coimbra, *Livros de Matrículas*) and, for the period 1772 -1820, from Manuel Alberto Carvalho Prata, *Ciência e Sociedade. A faculdade de Filosofia no período pombalino e pós -pombalino (1772-1820)*, Guarda, 1989". p. 5, n. 16.

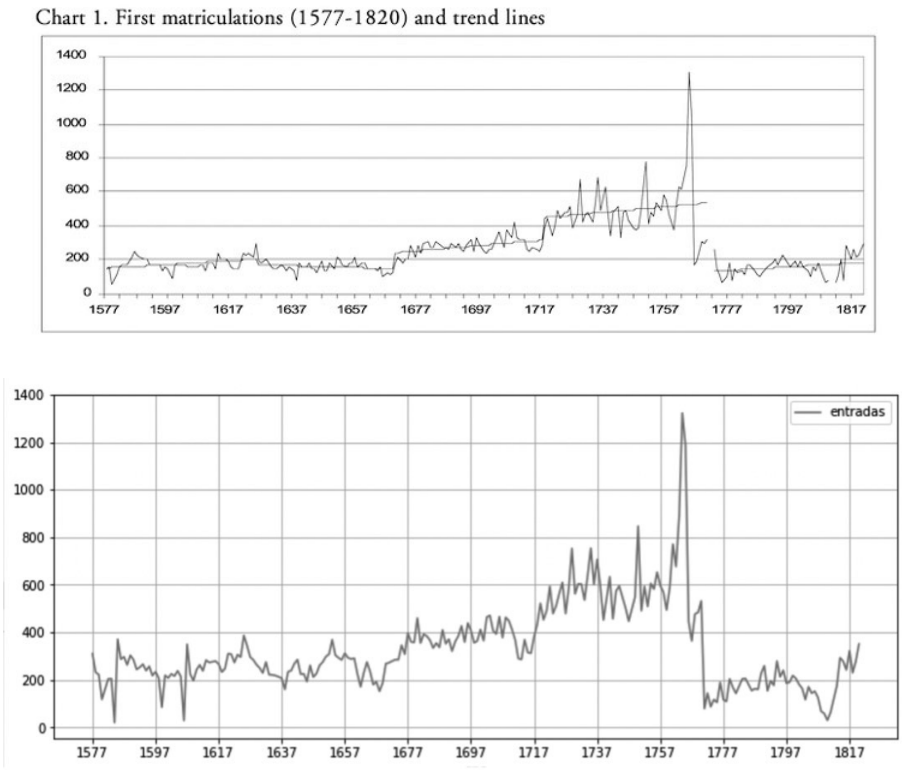


Figure 8 – Comparing data extracted from the sources according to Fonseca, 2007 (top), with data from the FA (bottom). There is a clear overall similarity in the chronological profile, with the FA showing higher numbers, especially in the 16th and 17th centuries, for reasons discussed in the text.

The first obvious conclusion is that the data extracted from the FA broadly maps the shape of data obtained from the "livros de matrículas" by Fonseca. The second conclusion is that the FA numbers are inflated, especially in the 16th and 17th centuries. This can be partially explained by the fact that Fonseca counts only the matriculations in Canon and Civil Law, the two main faculties in the number of students, which account on average, according to the author, to 87.3% of all enrolment. The role of duplicates is still an open question, to which there will be further references in this text.

Student population

Another opportunity for validation comes from the estimates of the student population between 1700 and 1770, also by Fonseca. Estimates of

the total population of students each year were calculated from the matriculations, cross-checked with a variety of sources[17].

To get comparable numbers, the fields "UnitDateInitial" and "UnitDateFinal" in the FA digital records were used. The fields contain the first and last date on the student record. It is assumed that the first date is the date of the first matriculation (as in the previous chart) and that the second date corresponds to the last year of "activity" of the student in the university.

For each year, the number of students with the first date in that year and with the last date in that year were totalled. The first number represents the year intake and the second number the students that will leave and will not be present in the following year, the outtake. With these two numbers it is possible to estimate the students present, without processing the yearly matriculations, which are complex to extract from the FA in a reliable way.

Again, the FA shows remarkable parallelism with data checked with the sources. And again, the numbers are higher than Fonseca's, for although this time both series cover all the faculties, they are computed differently.

Figure 9 – Comparing student population estimates. In the FA numbers a student is counted for every year between the first and final dates on the record, while Fonseca counts the actual matriculations of each year. Since some students do not enrol every year the FA numbers are naturally higher. Both series show the same general pattern.

Since we have the actual numbers for both series, we can go beyond the visual assessment and do a linear correlation.

17 *Cit.* note 15 *supra*, pp. 31-32.

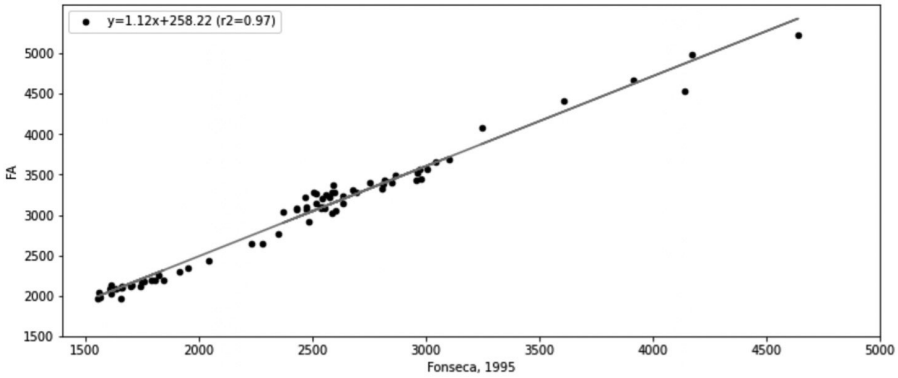


Figure 10 – Estimates of student population 1700-1770, FA vs Fonseca, 1995. On average, the numbers from FA are 23% higher. Considering the different methods of approaching the computation of the student population, the results are remarkably convergent.

Figure 10 shows the scatter plot for the two series. Each dot corresponds to a year, placed according to the values from Fonseca and the FA. The clustering along the regression line shows a strong relationship between the two estimates.

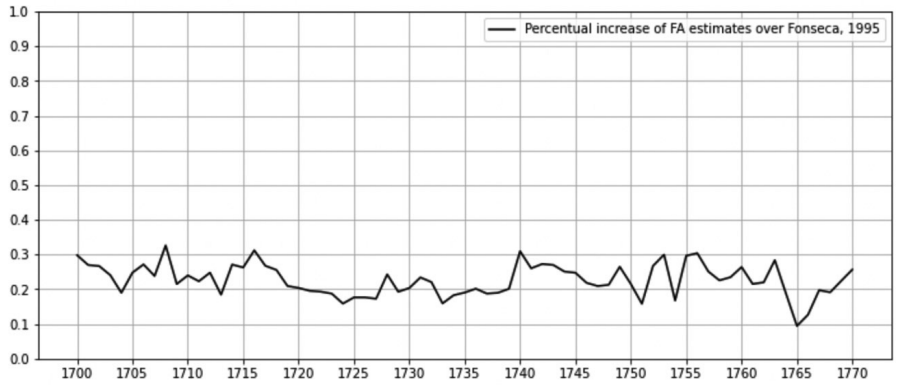


Figure 11 – Evolution over time of the percentual difference between estimates of the student population. There seems to be no obvious pattern over time. The average difference is 23% with a maximum of 33% and a minimum 0.9%, the latter being the outlier in 1765.

To control for the possibility that the divergence was concentrated in certain periods of time, the chronological evolution of the percentual difference over time was plotted. As can be seen in Figure 11, the variation fluctuates without a clear pattern, showing that the FA based estimate is consistently higher, in average 23%.

Note that the method used for computing the student population from the FA implicitly assumes that the students are present every year between the extreme dates on record. Of course, that is not the case, and there are several records that indicate long periods of absence between successive enrolments. So, in some way, the difference between the two estimates represents students that should have enrolled each year but did not and is a function of the combination of absenteeism and retention through suspension of study.

It is also possible that the process of production of the FA generated a steady flow of duplicates because first year students enrolled in pre-requisite studies and only further along their studies enrolled in their desired courses. This introduces a chance of starting a new record when a student first appears in the "faculdade" of graduation. This point will be addressed when analysing the problem of determining the "faculdade" from the FA records.

This section demonstrates that the data from the FA compares well in terms of chronological coverage with comparable data extracted directly from the sources. It also demonstrates that the question of duplicates is an open question, which can only be resolved by extracting as much information as possible from the records.

Personal attributes of students[18]

Annotations in the name

Student names can have annotations, between parenthesis. Currently 10,044 such annotations have been detected.

There are 839 variations of the expression used in the annotations. An effort was made to infer normalized values from that diversity of expressions and in this section the main sets of normalized attributes that can currently be extracted are presented and links for tables with nominal lists and statistical information are provided.

Example of a name with note (186633):

André Botelho (padre frei, colégio dos padres de São Pedro)

From this note, the algorithm can infer the following attributes:

18 Data for this section was generated with https://github.com/joaquimrcarvalho/fauc1537-1919/blob/main/notebooks/021-students_attributes.ipynb.

- Padre (priest)
- Title: Frei (friar)
- College: Colégio de S. Pedro da Ordem Terceira
- Religious order: Ordem Terceira

The most frequent attributes appearing in name notes, in their original forms, are shown in Table 3[19]:

Name note	N	First	Last
padre	5714	1537-00-00	1899-10-02
D.	600	1537-00-00	1910-10-20
frei monge de São Bento	174	1728-10-01	1799-10-29
frei monge de São Bernardo	154	1732-10-01	1827-10-29
padre frei religioso de Nossa Senhora do Carmo	96	1664-10-01	1760-10-01
padre frei religioso de São Bernardo	93	1664-10-01	1751-07-17
frei monge de São Jerónimo	90	1729-11-07	1825-10-25
padre frei religioso de Nossa Senhora da Graça	85	1669-10-01	1759-10-01
frei	71	1537-12-28	1830-10-22
padre frei religioso de São Bento	68	1660-12-17	1751-10-01

Table 3 – Most frequent expressions that appear between parenthesis after the names of students. The dates correspond to the date of first matriculation of the students, e.g., the first student with a note "D." (Dom) entered in 1537 and the last on 20th October 1910.

The notes in the original form are not very useful. Qualifiers such as "padre" (priest) and "frei" (friar) can appear in isolated form or as part of longer expressions. An attempt to normalize the note information was made by detecting the following specific information:

- Priesthood condition.
- Title understood as a qualifier of the person, including the "dom" and specific noble titles, but also religious titles such as "Bispo", "Cónego", "Prebítero", "Beneficiado, "Frei".
- College.
- Religious orders.

19 The full table is available at https://github.com/joaquimrcarvalho/fauc1537-1919/blob/main/inferences/name-notes/note_originals.csv

It would not fit in the scope of this paper to describe at length each of these personal attributes. Instead, listings for the most relevant attributes in the public repository are produced and the main ones are discussed[20].

Priests

In total, 6916 records with "padre" in the name note were extracted, the most frequent attribute in the notes. As with other attributes discussed in this section, it is possible to generate nominal lists with other attributes like place of birth, faculty, religious order, etc., as shown in Table 4[21]. From tables such as these, several interesting insights can be produced.

	Name	Entry	Last date	Place of birth	father	Order	Faculty	Title	College	Name note
150020	Manuel Batista	1564-10-01	1568-06-21				Teologia		Colégio de S.João Evangelista	padre do colégio dos Lóios
193400	Jorge	1564-10-01	1565-06-06			Ordem de Cristo	Teologia	frei	Colégio de Tomar	padre frei do colégio de Cristo
205601	António da Trindade	1578-06-30	1588-04-02			Ordem da Santíssima Trindade	Teologia	frei	Colégio da Trindade	padre frei do colégio da Trindade
201701	João da Costa	1580-03-22	1590-01-05	Lisboa	Fernão Nunes da Costa		Cânones	cónego		padre cónego na Sé de Coimbra
193431	Aleixo Jorge	1582-10-22	1594-06-11	Vila Boa	Jorge Gonçalves		Teologia	capelão		padre capelão da Universidade

Table 4 – Sample of students with "Padre" in the name note. Note that the algorithm can extract several attributes from a single note and link them with academic information such as "faculdade". In the first record, the current algorithm does not infer the religious order from the college because of the absence of the title "frei".

20 The tables for religious orders (3437 records), titles (4219) and colleges (379) can be found in the directory "inferences/name-notes" of the repository.

21 Full table at https://github.com/joaquimrcarvalho/fauc1537-1919/blob/main/inferences/name-notes/priests.csv

Religious orders

The religious order of the student is the second most frequent information that can be extracted from the name notes, with 3032 occurrences. There is a significant variation in vocabulary used to identify some of the orders; however, the algorithm maps expressions to religious orders with some success[22].

The college of residence was also used to infer the religious order. For instance, the college of Saint Thomas was the college of Dominicans in Coimbra, and students with the note "frei religioso do colégio de São Tomás" will be associated with the "Ordem de São Domingos".

The Augustinians are specially complicated because there were different branches in Coimbra. An annotation on record 217564 illustrates the problem, enumerating the different forms the student is referred to "... Frei Domingos de Santo Agostinho, Frei Domingos Eremita de Santo Agostinho ou Frei Domingos do Colégio da Graça"[23].

Through trial and error, it was possible to normalize this variety of notes into the list of religious orders in table Table 5.

Religious order	N	First	Last
Ordem de São Bernardo	401	1569-00-00	1827-10-29
Ordem de São Bento	374	1548-01-01	1831-06-25
Ordem do Carmo	283	1536-00-00	1829-10-31
Ordem de São Francisco	267	1540-03-09	1816-10-31
Ordem de Santo Agostinho (Graça)	250	1560-10-01	1780-05-30
Ordem da Santíssima Trindade	223	1551-10-00	1820-06-19
Ordem de São Domingos	202	1541-10-20	1827-10-30
Ordem de Santo Agostinho	168	1541-11-29	1829-06-03
Ordem de São Jerónimo	162	1550-03-12	1825-10-25
Ordem de Cristo	146	1563-10-01	1815-10-16
Ordem de Santo Agostinho (Descalços)	83	1734-10-01	1805-10-15
Ordens militares	71	1594-10-08	1815-10-20
Ordem Terceira	71	1632-00-00	1827-10-19
Ordem de São João de Deus	38	1575-10-01	1804-10-09
Ordem de São João Evangelista	36	1575-10-01	1769-10-01

22 The precise mapping between words in notes and religious orders is at https://github.com/joaquimrcarvalho/fauc1537-1919/blob/f033bc9f8bc9faad42f1ae3147256ac49fe39dba/notebooks/ucalumni/extractors.py#L74-L131

23 For the Augustinians see (Azevedo, 2011) For the colleges of the religious orders the reference is "Os Colégios universitários de Coimbra" (Vasconcelos, 1987, vol. I, pp. 155-295).

Religious order	N	First	Last
Ordem de São Paulo	35	1670-10-01	1829-10-24
Ordem dos Cónegos Regrantes de Santo Agostinho	25	1543-12-21	1825-05-26
Ordem de São Pedro	16	1670-00-00	1805-11-23
Companhia de Jesus	13	1555-06-06	1557-02-20

Table 5 – List of religious orders with more than 10 occurrences. "Ordens militares" refers to an unspecified military order, except the "Ordem de Cristo". The "Ordem de Cristo" had its own college in Coimbra, while all the others shared the same college, normally referred as "Colégio dos militares".

Around 70 expressions have escaped normalization so far, many because of misspellings. A nominal list of all the students with a recognized religious order is provided in the repository[24].

An interesting detail of Table 5 is the 13 Jesuits between 1555 and 1557. They are certainly related to the transfer of the "Colégio das Artes" to the Society of Jesus in 1555[25]. This created the need to award degrees recognized by the university to its teachers. In the short span of their recorded university life, the Jesuits in the FA followed a fast track to a bachelor's degree in Arts Among those few are famous names such Luís de Molina, Nicolau de Gracida and Sebastião de Morais[26].

Listings for the most relevant attributes have been produced and these are available in the public repository.

Place of birth

Place of birth is the most common personal attribute in the FA, with 98902 occurrences.

The list of places occurring more than 500 times is shown in Table 6.

24 https://github.com/joaquimrcarvalho/fauc1537-1919/blob/main/inferences/name-notes/religious-orders.csv.

25 The "Colégio das Artes" provided the preparatory studies required to attend university degrees, the equivalent of modern secondary education. In the aftermath of a scandalous process involving the Inquisition and teachers of the college, the king reached an agreement with Jesuits for them to take responsibility for the college and provide the teaching staff.

26 The issues related to the award of university degrees to the Jesuit teachers of the "Colégio das Artes" are described in Brandão e Almeida (1937, vol. I, pp. 220-224). It is interesting that conjunctural episodes leave their imprint in the FA, assuming those 13 name notes were not the result of a *posteriori* interests that found their way into the catalogue.

Place	N
Lisboa	9484
Coimbra	5785
Porto	3591
Braga	1699
Évora	1160
Guimarães	1051
Lamego	1033
Viseu	1031
Aveiro	844
Vila Real	824
Santarém	791
Leiria	697
Ilha da Madeira	623
Portalegre	579
Tomar	568
Viana	566
Castelo Branco	562
Barcelos	552
Beja	532
Baía, Brasil	527
Guarda	519

Table 6 – List of the most common places of birth.

The value of this list is impacted by the inevitable misspellings and by the lack of uniformity in providing a wider geographic or administrative context when needed. For instance, Ponta Delgada, the main city in the Azores archipelago, is registered in eight different forms. Together, the islands of Azores and Madeira generate 142 different place names, and there is no way to query them in a structured way.

The convention of separating successive levels of geographic identity by commas, as in "Ponta Delgada, Ilha de São Miguel, Açores", although the most common, coexists with other forms of punctuation based on slashes, hyphens and parenthesis.

Standardization of the geographic names in the FA would greatly increase the applications possible with the structured data, especially mapping and spatial analysis. In the repository, a list of all values for place-of-birth is provided[27].

27 https://github.com/joaquimrcarvalho/fauc1537-1919/blob/main/inferences/places.

Parents

The names of the fathers and mothers were also extracted, albeit the latter is present only in less than a hundred cases. Currently there do not seem to be many usages for this information, except detection of brothers. Transgenerational identification would be interesting, to detected the linages that certainly exist, but is difficult to automate.

Academic information

Information related to personal attributes of the students has been dealt with, but the core information of interest of the FA lies in the academic information encapsulated in a single text field "ScopeContent". This field records the academic path of the student and can include enrolments, exams and degrees obtained.

Processing the academic information faces two major challenges. The information is recorded in a text field with many variations in notation and terminology. The second is substantive: the FAUC includes information spanning four centuries. Academic procedures, degree structures and record keeping methods varied greatly during that period. Processing the information about the faculty in which students enrolled illustrates these difficulties, due to the impact of changing degree structures, the way the academic path was differently recorded in the sources and how paths involving more than one faculty were interpreted.

Faculty

Determining the "Faculdade" is the most important step in processing academic information. Without a usable value for that essential element of academic identity, little interest can be done with the information from enrolments, degrees, and exams. More importantly, many errors can be introduced in the downstream analysis if the value of "Faculdade" is wrongly determined.

The field "Faculdade" is the first line of the section with academic information on the FA. Unfortunately, its contents are not adequate for several purposes. There are severe limitations that may not be very important for the human user locating a specific student, which can generate substantial distortions in the bulk data processing.

The main problems with the contents of the "Faculdade" field in the FA are:

- Missing value: no value in field (12,000).
- Inconsistent value: the field value is inconsistent with the rest of the academic information (7,500).
- Inconsistent chronology: the value corresponds to a faculty that did not exist at that date (500).
- Erroneous value: the content of the field does not correspond to a recognizable faculty, mainly due to misspellings or input of information belonging to other fields (700).

Altogether these problems amount to 20% of the records, and therefore considerable effort was invested in exploring to what extent they can be addressed in an automated way.

To understand why there are so many problems, especially of the two first types, it is necessary to understand the relation between the production of the FA and the original records.

The main sources for the FA are the "livros de matrículas", which, from 1577 onwards, provide the enrolment registry. At the start of each academic year, the students would arrive at Coimbra and enrol. Enrolment was compulsory and several mechanisms were deployed over the years to ensure enrolment on time to check that students remained in the city after signing their "matrícula".

Until 1834, most of the students aiming for a degree in Civil or Canon Law. The curriculum of the first two years of both degrees consisted of common introductory courses, namely the course "Instituta". So "Instituta" is the first enrolment on the records for the majority of first year students[28].

This first enrolment produces a card in the original FA. At that point, it is not clear if the student would continue later to Canon Law or Civil Law. As a result, the information on "Faculdade" is left either blank, with "Direito", or simply with "Instituta"[29]. Many records of the digital FA also have simply

28 The most important changes in academic structures that impact the understanding of the FAUC are the reforms of 1772 and 1834. The reference work on this matter is Fonseca (1995, chap. I) complemented by Fonseca (2000) for the fine details of academic regulations. A more precise understanding of the changing regulations can be obtained by analysing the various "Estatutos". See the references for the list of available printed sources.

29 The faculty of "Direito" is created after the reforms of 1834, by merging "Cânones" and "Leis". References to "Direito" as "Faculdade" before that date are to be understood as a generic reference to the law faculties and not to be confused with the real faculty of "Direito".

a question mark for the field "Faculdade" when the information is a single 'Instituta' enrolment.

As the student later enrols in the destination faculty there is a chance that a new record is created. See Figure 6 above as an example of such duplication.

The "Instituta" course disappeared in the profound reformation of 1772. However, a similar situation was created in the new study plans. All students, irrespective of the area of study, should do introductory courses at the Faculty of Mathematics and Philosophy for two years, and only then continue to the school where they would receive their degree. In this context, the first enrolments occur in the faculty of Mathematics or Philosophy despite the students going on to study Civil or Canon Law, Medicine, or Theology. Note that some students did go on to graduate in Mathematics or Philosophy as their first choice.

This means that, after 1772, the first enrolment of a new student in the university is done in a faculty which will necessarily be the one of graduation. These students would be known as "obrigados" to be distinguished from the "ordinários", those that really wanted to study mathematics and philosophy.

This situation affects badly the semantics of the field "faculdade" on the FA. In a unified catalogue, with one record per student, the content of the field "faculdade" should be the one in which the student graduates, taking care not to miss the double degrees that existed, mainly in Canon and Civil Law.

What we there is instead is a situation where cards were produced when the information about the main faculty was not available and were left with the field blank or, worse, with a value that was in fact a pre-requisite. Subsequent information was added to the cards without correcting the initial limitations. It is possible that different people, working either in the original catalogue or in the digital version, made different decisions in similar situations and interpreted differently the purpose of the field.

Nevertheless, since the subsequent information is in the record, it is possible to attempt to overcome those limitations.

The algorithm tries to infer the true faculty of every record by analysing all the information available. Using lists of pre-requisites in different periods, it was possible to infer the faculty in cases where it was absent and correct cases where the faculty on the record was inconsistent with the rest of the academic information.

Instead of going into a discussion of decision logic, an example of what is possible to automate might be useful. The case in Figure 12 is interesting because it is complex, and the result can be externally validated. On the left of Figure 12 is the FA record and, on the right, the inferred academic path.

FA Record	Algorithmic reconstruction of academic path		
Âmbito e conteúdo Faculdade: Matemática Matrícula(s): 19.10.1788 (obrigado) 03.10.1794 Direito: 31.10.1794 02.10.1795 Leis: 03.10.1796 04.10.1797 02.10.1798 09.10.1799 Filosofia: 31.10.1793 06.10.1795 Exames: 3.º 08.07.1797 Aprovado Nemine Discrepante, Atos n.º 6, fl. 19v. 4.º e grau de Bacharel 20.06.1798 Aprovado Nemine Discrepante, Atos n.º 6, fl. 95v. Formatura 08.06.1799 Aprovado Nemine Discrepante, Atos n.º 6, fl. 186 2.ª Formatura 03.07.Aprovado Nemine Discrepante Repetição 22.06.1801, Atos grandes n.º 2, fl. 10 Exame privado e grau de Licenciado 06.07.1801, Atos grandes n.º 2, fl. 104 Doutoramento 19.07.1801, Atos grandes n.º 2, fl. 182v. Graduação em Direito 22.10.1800	**Date**	**type**	**value**
	1788-10-19	faculdade	Leis
	1788-10-19	faculdade-original	Matemática
	1788-10-19	matricula-faculdade	Matemática
	1793-10-31	matricula-faculdade	Filosofia
	1794-10-03	matricula-faculdade	Matemática
	1794-10-31	matricula-faculdade	Direito
	1795-10-02	matricula-faculdade	Direito
	1795-10-06	matricula-faculdade	Filosofia
	1796-10-03	matricula-faculdade	Leis
	1797-10-04	matricula-faculdade	Leis
	1798-06-20	grau	Bacharel em Leis
	1798-10-02	matricula-faculdade	Leis
	1799-06-08	grau	Formatura em Leis
	1799-10-09	matricula-faculdade	Leis
	1801-07-06	grau	Licenciado em Leis
	1801-07-19	grau	Doutor em Leis
	2022-05-01	nota-processamento	Aviso: faculdade corrigida

Figure 12 – Record 215786, João da Rocha Dantas e Mendonça. By considering the pre-requisite rules at the time, the algorithm corrects the field "Faculdade" in the FA and disambiguates both enrolments and degrees awarded. The inferred doctoral degree in "Leis" can be confirmed in Vasconcelos (1987, vol. II, p. 41).

The FA registers faculty as "Matemática" based on the first record on the books of the student as "obrigado" in 1788, meaning it was a pre-requisite enrolment. The FA then notes, six years later, the enrolment in "Direito", an expression used for the introductory law courses for students of "Cânones" and "Leis". Only in 1796, the student enrols in the destination faculty, "Leis". The enrolment in the pre-requisite "Filosofia" is also recorded.

On the right is the academic path, as inferred by the algorithm. The presence of enrolments in "Leis", triggers the replacement of "Faculdade: Matemática" with "Faculdade: Leis". But notice that the two first "matrículas" are correctly attributed to "Matemática" because unspecified enrolments are attributed to the original "faculdade" in the FA record. The process manages to correctly attribute the successive dates of the "matrículas" section, based on the presence of faculty names. Also, the inference allows the process to correctly identify the area in which the degrees were obtained, which is crucial. The result is a clear path that abides by the academic statutes of the time, even if it misses such details as the repetition of "Formatura" and the "Graduação" in 1800 and keeps, for tracing purposes, the expression "Direito" in the enrolment, when "Cursos Jurídicos Comuns" or something similar would be more appropriate.

Incidentally, note the number of years that elapse between the first and the last date on record, and that there are no enrolments between 1788 and 1793. This type of record is responsible for the difference in estimating the student population from start and end dates, instead of adding up enrolments, as was mentioned in the section on coverage above.

In narrative terms, the algorithm checks first if the field "Faculdade" has a value or not; it then proceeds to collect all references to "faculdades" in the rest of the record. In the next step, it removes all the pre-requisites from the collection: if something remains it will the chosen as the corrected "faculdade"; if nothing remains, then the original value is kept[30].

The same process allows us to "infer" the "faculdade" when the field in the original record is empty and to flag as errors values that are not in the list of schools in each period, and uncommon combinations (Table 7).

Type of changes to "Faculdade"	**N**
Value inferred (original empty)	11356
Value corrected (original inconsistent)	6980
Uncommon academic path	735

Table 7 – Algorithmic changes to the value of field "Faculdade".

When attempting to infer the "Faculdade" it is not always possible to solve the ambiguity between "Cânones" and "Leis". This is due mainly to many records containing only the registration in "Instituta", but also because the field "Faculdade" was not updated when further information was collected. There are records where it is possible to follow the academic career of a student up to the doctorate level without solving the ambiguity between "Cânones" and "Leis" because the ambiguous expression of "Direito" was never updated[31]. Half of the inferred values for "Faculdade" are still ambiguous.

Inferred values for Faculdade	
Cursos jurídicos (Cânones ou Leis)	6993
Artes	3005
Cânones	764
Leis	386

30 The algorithm for "Faculdade" is 150 lines long and the version at the time of writing is at https://github.com/joaquimrcarvalho/fauc1537-1919/blob/f033bc9f8bc9faad42f1ae3147256ac49fe39dba/notebooks/ucalumni/extractors.py#L460-L611

31 See, for instance, 206171, Jerónimo José Rodrigues.

Inferred values for Faculdade	
Teologia	352
Medicina	185
Filosofia	28
Matemática	24
Direito	22

Table 8 – Inferred values for empty "Faculdade" field. In more than half the cases it is not possible to determine from the records in which of the law schools the student was enrolled.

Correcting an existing expression is more delicate that inferring missing information because it involves overriding human input. The algorithm currently suggests close to 7,000 changes. These include either changing the name of the faculty or adding a second name to the existing one, assuming a double degree path or a change of main "faculdade".

The most frequent correspond to what would be expected, but they are also some other combinations that are flagged as "uncommon", such as Theology with Law and Medicine with Canon Law that require further analysis[32].

Changes suggested algorithmically	N	First	Last
Cânones → Cânones,Leis.	1254	1536-10-00	1821-11-12
Leis → Cânones,Leis.	1246	1537-00-00	1829-06-13
Matemática → Medicina.	1175	1772-11-28	1909-11-11
Direito → Leis.	1023	1687-01-05	1827-06-16
Filosofia → Medicina.	480	1540-10-25	1909-11-09

Table 9 – Main changes suggested by the processing of "Faculdade". The most common are related to double degree paths in Canon and Civil Law, the replacement of pre-requisite studies in Mathematics and Philosophy by Medicine and the disambiguation of "Direito" into "Leis" or "Cânones".

The problems related to the field "Faculdade" can be overcome by inferring missing values and producing suggestions of correction of existing values based on information from the rest of the record.

A more fundamental problem is the definition of what should go into the field "Faculdade". There are records where all possibilities found on the record are enumerated in the field, separated by slashes. This indicates a

32 For instance, in record 127769, José Cabaço, the suggestion is made to change "Cânones" to "Cânones, Leis", because there is a reference to "Formatura em Leis".

fluctuation of the meaning of the field, which is natural due to the history of the FA, but should be avoided[33].

Degrees

Extracting degree information from the FA is relatively simple, because it can be triggered by a very limited set of well-defined terms. Most of the references to degrees do not specify the area, which means that the prior processing of "Faculdade" is necessary. Here are the degrees with over 500 occurrences.

Degree	N
Bacharel em Cânones	21744
Formatura em Cânones	18823
Bacharel em Artes	6342
Bacharel em Leis	5963
Formatura em Leis	5066
Bacharel em Direito	4848
Formatura em Direito	4597
Licenciado em Artes	2834
Formatura em Medicina	2299
Bacharel em Cursos jurídicos (Cânones ou Leis)	2217
Licenciado em Cânones	2131
Formatura em Cursos jurídicos (Cânones ou Leis)	1437
Licenciado em Medicina	1349
Licenciado em Teologia	1127
Doutor em Cânones	965
Formatura em Teologia	834
Bacharel em Teologia	834
Doutor em Teologia	765
Licenciado em Leis	640

Table 10 – Most frequent degrees (with more than 500 students).

The ambiguity in the Law degrees is inherited from the faculty, as explained in the preceding section. In fact, except for "Artes", it is not very frequent that the record explicitly states the area of the degree. The algorithm will use any explicit reference to the faculty in the degree expression,

33 The record 202402, Vicente Coelho da Silva Seabra e Teles is an example with "Faculdade: Matemática / Filosofia / Medicina".

and if none is found, uses the value for "Faculdade", explicit, inferred or corrected[34].

Enrolment ("matrículas")

Enrolments make up most of the academic information in the records. Unfortunately, the detail of the recording varies greatly. The rules of pre-requisites to referred above implied that various types of qualifications were added, denoting different "faculdades" and different modes of enrolment ("obrigado", "ordinário", "voluntário"). Additionally, in recent years, enrolment information was added with the curricular year and precise references to the original sources. This rich information, easily absorbed by a human mind, poses many challenges to automated processing.

Currently, the algorithm tries to extract different attributes of each enrolment:

- The scope: faculty (explicit, inferred or corrected), year in the study plan, specific course.
- The mode: "ordinário", "obrigado" or "voluntário".
- The date: a great variety of date formats are recognized.
- Sources: references to books and page numbers in the original sources.

The "Provas" or proofs, a mechanism to assert that a student had attended the necessary classes before the enrolment system was established, are not currently processed. The student would come to the registrar with two witnesses to declare that the required years of study were fulfilled. In the FA this is recorded with the expression "Provou", which is easy to detect, but it is hard to process the rest of the sentence.

Exams

Processing the exam information that has been added recently to the FA is still a work in progress.

34 We provide the list of students with a doctoral degree, in all areas; https://github.com/joaquimrcarvalho/fauc1537-1919/blob/main/inferences/academic/doutores.csv

The example in the Appendix shows information of exams correctly recognized by the algorithm, but more fine-tuning is necessary to handle the many variations found in the records. The problem is that no established style exists for important details such as repetition of failed exams.

Moreover, the algorithm does not currently have a representation of which exams were required for each study plan in different periods. As a consequence, it cannot have a "pull" strategy based on a set terminology, like with degrees, and tries to take whatever has the word exam.

Currently, there are close to 54,000 references to exams, most for the period after 1772, so it is worth continuing to perfect the extraction.

Summary of current extraction capabilities

Table 11 lists the attributes that, at the time of writing, are extracted from FA records and introduced in a database from which tables and lists such as those used here were produced.

Attribute	Number of students	Note
colegio	356	
código-de-referência	105298	ISAD ref. code
data-do-registo	105298	Last change of record
exame	53697	
faculdade	104638	Final after checks
faculdade-original	7167	Original value on record
grau	87112	
instituta	40672	
matricula-classe	2444	Year of study plan
matricula-curso	170	Specific course
matricula-faculdade	313032	
matricula-outra	1373	Scope not understood
matricula-universidade	571	No information on scope
naturalidade	98902	
nome	113923	
nome-apelido	198380	
nome-geografico	107415	Segment of naturalidade
nome-mae	93	
nome-nota	10044	Original name note
nome-pai	49916	
nome-primeiro	105298	First name
nome-vide	8625	Cross-ref. expression

Attribute	Number of students	Note
nota-processamento	20263	Errors and warnings
ordem-religiosa	3081	
padre	7059	
titulo	4207	D., friar, nobility
uc-entrada	105298	
uc-saida	105298	
url	105298	Link to FA record

Table 11 – List of attributes extracted from the FA records. See Appendix for a specific example.

Evolution of the digital FA, from "cathedral" to "bazaar"

The previous sections demonstrate that it is possible to generate structured information from existing FA records. Although much can still be improved in terms of what an automated process can do, relevant progress is dependent on two key developments: first, increase the quality of the FA through normalization of the information in free-text fields; secondly, correct mistakes and solve ambiguities that the records contain. The last aspect relates to the second aim of this text, which is to propose an approach to the evolution of the FA that incorporates the lessons learned from the open-source model. The argument is that the evolution of the FA should harvest the enormous amount of dispersed knowledge about the students of the University of Coimbra that exists in various forms, by encouraging the development of a community of "contributors" along with the model of open-source projects, as a complement to the centralized institutional efforts (a "bazaar" around the "cathedral", as explained below). Furthermore, the evolution of the current tools in the data science field and the growing amount of linkable data provide opportunities that should be taken into consideration. The next section argues for an "Open Data Science" approach for the future of the digital FA.

The evolving tools of Data Science

The results presented here were made possible by the availability of computer tools associated with the emergent field known as "Data Science". The definition in Wikipedia is particularly adequate here: "Data science is an interdis-

ciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from noisy, structured and unstructured data"[35].

The vast amount of data collected and made accessible on the Internet, together with the importance for many industries of the unstructured information produced daily by millions of social media users, stimulated the production of new computer tools tuned for "noisy structured and unstructured data".

The chronic shortage of expertise in computer science further ensured that the tools evolved to be userfriendly and easy to master so that a "data scientist" is potentially a larger demographic than the traditional computer programmer. Although most practitioners have a background in Computer Science, Mathematics or Statistics there is no lack of encouragement for people of the most diverse areas to join the new discipline[36].

Even more significantly, big software corporations have invested heavily in a new generation of easy-to-use tools. Visual Studio Code by Microsoft is becoming the Microsoft Word for data science projects, allowing users to write text intermixed with data processing instructions, graphs and maps, in the form of "notebooks" that can be shared and reused by others[37].

The trend also includes a high degree of standardization of data representation formats. The popularity of the Pandas data analysis software, specially designed to coexist with "notebooks", has eclipsed more traditional statistical software [38]. Once one's data is imported into a Panda "dataframe", it is a short path for obtaining statistics, graphs, maps, and network analysis. The production of something such as the section above on coverage is relatively straightforward.

Open Science as Open-Source

Another development of relevance to the future of the FA is the popularization of collaborative tools initially developed for open-source coding projects. These started as "version control systems" that allowed several

35 (*Data science*, 2022). The recent multiplication of bachelor's and master's degrees on data science in Portugal is a testimonial of the mainstreaming of a field that until recently was a "community of practice" sharing a set of new software tools. For an example see *Licenciatura em Engenharia e Ciência de Dados - Departamento de Engenharia Informática – Cursos – Universidade de Coimbra*, s.d.

36 (Kim, 2021).

37 (Jupyter Notebooks in Visual Studio Code, n.d.).

38 (*Pandas – Python Data Analysis Library*, n.d.).

people to cooperate in producing computer software, facilitating the merging of different versions while tracing precisely who did what and when. These early systems evolved into environments where collective text-based cooperation happens spontaneously between people that are complete strangers and under no formal authority.

In a famous, albeit somewhat dated book, "The Cathedral and the Bazaar", Eric Raymond described the impact of the early "version control systems" in how some of the most complex computer software was produced, replacing the highly planned, hierarchically organized model of the cathedral by the more spontaneous, self-organized and ultimately more successful way of the bazaar[39].

What was an argument for the "hacker" community in 1999, has become the dominant model today. The acquisition in 2018 by Microsoft of GitHub, the biggest host of open-source projects, increased a trend already underway to spread the usage of open-source collaborative tools by a wider community. The integration of GitHub in Visual Studio code makes collaborative data science an accessible reality. The great classroom and helpdesk of the Internet, with its abundance of high-quality tutorials, answers to frequent questions, recipes for all types of data manipulation and visualization, contributes enormously to lowering the difficulties of adoption of the new approaches to data processing and collaboration.

The FA as an Open Data Science Opportunity

The FA is an amazing data set. The academic information, for a period of near 400 centuries, of the intellectual elites of a country whose language and culture has a global presence, is no common database. The accumulated knowledge of the FA, the result of generations of patient and highly competent archivists and historians, is a treasure that still has many untapped gems in this new age of data science, collaborative tools and linked data.

Maintaining a single source of truth

The exploratory work presented here was often faced with a difficult trade-off: automated correction of problems vs correction of the original records.

39 (Raymond, 1999).

There is an irrefutable argument in favour of correcting the original records: they constitute the necessary "single source of truth" about the students of the University of Coimbra[40]. But, at the same time, thousands of corrections can be done downstream from the original FA and enable many interesting analyses that can contribute to the improvement of the catalogue and more efficient usage of the human resources allocated to it.

Additionally, the milestone publication online of the FA, with known limitations, was done to stimulate further research by people who would not easily come to the archive to search the old paper catalogue. It certainly enabled the detection of errors by the large community that uses it as a starting point for studies of all types, from family biographies to postgraduate theses and history books.

Managing contributions

The argument is that the model for managing a large collaborative effort while maintaining a reference "implementation" already exists in the open-source environment. The counterargument that those tools are too technical and difficult is losing ground; all well-managed open-source projects have a clear policy on "how to contribute" that starts normally with the very non-technical but crucial tasks of reporting "bugs", revising the manuals and producing tutorials. On top of the increased ease of use of those tools, there is also a hierarchy of contributions that can be defined in our context.

It is suggested that the standard open-source concept of "issues" be used to harvest contributions. An issue is a report of a problem in software, data or documentation. An issue starts with a description of the problem, normally in a predefined form, and initiates a discussion about how to solve it until a new version is released that "closes" the issue.

The typology in the case of the FA is not difficult to define:

- Problems that do not require examining in the original cards: misspellings, and punctuation issues that impact data extraction.
- Problems that require checking with the paper original, such as place of birth registered in the field "faculdade", or badly formed dates.

40 (Single source of truth, 2022).

- Problems that require revisiting the sources, such as records with missing information about "faculdade" but with references to the original books where the information should be available.
- Problems that are detected because of information in other sources, external to the university archive, for instance, a student with place of birth "Loure" was detected who was in fact born in "Soure". The two words can be impossible to distinguish in manuscript records, but the mistake is easily detected by someone who happens to have researched the person in question in other archives. Many errors of this type could be reported by many users of the FA.
- Normalization, especially of place names.

A platform like GitHub has easy to use interfaces and guidelines for the management of issues, both from the point of view of the reporter as well as of the people responsible for the corrections[41]. The record ID provides an easy way to tag the issue and keep track of the problems of specific records. In addition, issues can be prioritized for better planning with the tools available. Using a public open-source repository where various listings are made available and support documents for normalization are shared is an important complement.

Recommendations

To prepare the FA for a new life in the Open Data Science world there are a few simple things that could be done.

1. Keep the punctuation of the academic record consistent by following the most common format of "field: value".
2. Adopt a general structure for the record, based on the most common format and apply it consistently. The best records to process have a section/field structure, with "Matricula(s):" introducing a part with dates preceded by names of school, or year of study. The same should be used for exams and degrees.
3. Clearly define the semantics of the field "Faculdade". Our suggestion is that the field should correspond to the faculty of graduation.

41 (*About issues*, n.d.).

Cases of double degrees, such as Canon and Civil Law, should produce two values in the field, separated by a slash.

4. Minimize the uncertainty of the field "Faculdade". When the field has no value or "direito" was used as a placeholder for Civil or Canon Law, and the record is being updated for some reason, it is worth the extra effort to disambiguate "faculdade".
5. Ensure that each student has a stable and simple identification number, such as the six-digit number that the current software provides in the URL of the web records. The existence of a stable "id" for each student is a basic requirement to take advantage of linked data opportunities in the future.

Other not so simple recommendations relate to cross-references and the normalization of place names.

1. Explore the possibility of migrating the cross-reference information to authority records as defined by the ISAAR standard and implemented in the *Archeevo* software.
2. Likewise, explore the possibility of normalizing the list of placenames that can be easily extracted from the records and processed for variations.
3. Consider enabling the OAI-PMH interface to the FA, so that undertakings such as the one presented here are not dependent on direct exports from the archival management software in use[42].

Conclusion

There is much more about the structuring of the information of the FA and the possibilities opened by the new world of data science that could not find space in this paper.

This paper hopes to demonstrate that it is possible to structure information collected in cards decades ago and bring them to the modern world of data science tools.

The repository associated with this paper will provide further information of interest to those who would like to take the challenge of improving a data set which is a core element of the intangible heritage of an institution that deserved the UNESCO World Heritage listing.

42 (*Open Archives Initiative Protocol for Metadata Harvesting*, n.d.).

Acknowledgements

I wish to thank the Director of the Archives of the University of Coimbra, Cristina Freitas, for supporting the idea behind this project and facilitating an export file of the FA. Ana Maria Bandeira, the archivist in charge of the university collection, was an invaluable source of insights and expert knowledge about the FA and the sources behind it. I join with the gratitude of numerous researchers and students who like me always found a welcoming smile and wise from her. Catarina Macedo helped in overcoming the limitations of distance in a period when travelling to Coimbra was impossible, helping in the validation of results and procuring photographs of the original cards when a fine point needed to be understood. David Sansom generously reviewed the manuscript and made many suggestions that greatly improved the final text. Finally, I am much indebted to Macao Polytechnic University for providing the working environment that made this work possible.

Bibliography

Archeevo – Arquivo da Universidade de Coimbra. (2021). *Índice de alunos da Universidade de Coimbra*. https://pesquisa.auc.uc.pt/details?id=264605

Azevedo, C. A. M. (2011). *Ordem dos Eremitas de Santo Agostinho em Portugal: 1256-1834, História religiosa* (1ª ed.). Centro de Estudos de História Religiosa.

Bandeira, A. M. (s.d.). *Percurso académico na Universidade de Coimbra, nos séculos XVI a XX (orientações para pesquisa)*. Arquivo da Universidade de Coimbra. https://www.uc.pt/auc/orientacoes/UC_GuiaPercursoAcademico.pdf

Brandão, M., & Almeida, M. L de. (1937). *A Universidade de Coimbra: esbôço da sua história* (1ª ed.). Por Ordem da Universidade.

Fonseca, F. T. da. (1995). *A Universidade de Coimbra 1700-1771: estudo social e económico, Acta Universitatis Conimbrigensis* (1ª ed.). Por Ordem da Universidade.

Fonseca, F. T. da. (2000). A dimensão pedagógica da reforma de 1772: alguns aspectos. In A. C. Araújo (Coord.), *O Marquês de Pombal e a Universidade* (1ª ed.) (pp. 43-68). Imprensa da Universidade de Coimbra. https://digitalis.uc.pt/handle/10316.2/32748

Fonseca, F. T. da. (2007). The social and cultural roles of the University of Coimbra (1537-1820): some considerations. *e-journal of Portuguese History, 5*(1).

Ghdocs – prod.azurewebsites. (n.d.). *About issues*. https://ghdocs-prod.azurewebsites.net/en/issues/tracking-your-work-with-issues/about-issues

International Council on Archives Conseil International des Archives. (2000). *ISAD(G): General international standard archival description; adopted by the Committee on Descriptive Standards, Stockholm, Sweden, 19-22 Sept. 1999. ICA Standards* (2nd ed.). International Council of Archives.

Keep Solutions. (2022). *Archeevo – Archival management software for the semi-active and inactive stages of information*. https://www.keep.pt/en/produts/archeevo-archival-management-software/

Kim, D. (2021). *Transitioning from Social Science to Data Science* (1st ed.). https://towardsdatascience.com/transitioning-from-social-science-to-data-science-7e22501b5a3b

Open Archives. (n.d.). *Open Archives Initiative Protocol for Metadata Harvesting*. https://www.openarchives.org/pmh/

Pandas – Python Data Analysis Library. (n.d.). https://pandas.pydata.org/

Raymond, E. S. (1999). *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary (O'Reilly Linux)* (1st ed.). O'Reilly.

Universidade de Coimbra. (2020). *Cursos / Departamento de Engenharia Informática – Licenciatura em Engenharia e Ciência de Dados*. https://apps.uc.pt/courses/PT/course/8502

Vasconcelos, A. de. (1987). *Escritos vários* (reed.). Arquivo da Universidade de Coimbra.

Visual Studio Code. (2022). *Jupyter Notebooks*. https://code.visualstudio.com/learn/educators/notebooks

Wikipédia. (2022). *Data science*. https://en.wikipedia.org/w/index.php?title=Data_science&oldid=1087062839

Wikipédia. (2022). *Single source of truth*. https://en.wikipedia.org/w/index.php?title=Single_source_of_truth&oldid=1084162596

Sources for academic structures, curricula, and regulations

Estatutos da Vniuersidade de Coimbra confirmados por el rey Dom Phelippe primeiro deste nome, nosso Senhor em o anno de 1591-. (1593). Em Comibra [sic]: por Antonio de Barreira.

Estatutos Da Universidade De Coimbra: Compilados Debaixo Da Immediata E Suprema Inspecção De El Rei D. José I. Nosso Senhor Pela Junta De Providencia Literaria ... Deste Presente Anno de 1772. (1773). Regia Officina Typografica.

Leite, S. (1963). *Estatutos da Universidade de Coimbra (1559), Acta universitatis conimbrigensis* (1ª ed.). Por ordem da Universidade.

Rodrigues, M. A., & Arquivo da Universidade de Coimbra (1991). *Universidade De Coimbra. Os primeiros estatutos da Universidade de Coimbra* (1ª ed.). Arquivo da Universidade de Coimbra.

Appendix

Example of the current capabilities of the algorithm including correction of "Faculdade", extraction of religious order, titles, different scope of enrolments, exam results and degrees. Some limitations in interpreting enrollments can be seen, regarding the two unqualified dates.

José Doutél (frei monge de São Bernardo)

Nível de descrição

Documento simples

Código de referência

PT/AUC/ELU/UC-AUC/B/001-001/D/001409

Tipo de título

original

Título

José Doutél (frei monge de São Bernardo)

Datas de produção

1798-10-15 a 1807-01-15

História administrativa/biográfica/familiar

Filiação: António Venceslau Doutel
Naturalidade: Rio Bom

Âmbito e conteúdo

Faculdade: Matemática
Matrícula(s): 15.10.1798 (obrigado)
2º ano - 02.10.1799
Filosofia - 15.10.1798
2º ano - 03.10.1799
3º ano - 03.10.1800
09.10.1802
22.10.1803
Teologia - 20.10.1800 (ordinário)
2º ano - 23.10.1801
3º ano - 15.10.1802
4º ano - 03.10.1803
5º ano - 06.10.1804
Graduação - 31.10.1805
Exames: 1º ano: 11.06.1801, Aprovado Nemine Discrepante, Atos n.º 3, fl. 5
2º: 18.06.1802, Aprovado Nemine Discrepante, Atos n.º 3, fl. 34
3.º exame: 08.06.1803, Aprovado Nemine Discrepante, Atos e Grau n.º 4, fl. 55
4º exame e Grau de Bacharel: 13.06.1804, Aprovado Nemine Discrepante, Atos e Grau n.º 4, fl. 96
5º exame: 06.06.1805, Aprovado Nemine Discrepante, Atos n.º 4, fl. 135
Exame Privado e Grau de Licenciado: 18.07.1806, Atos Grandes n.º 3, fl. 26
Grau de Doutoramento: 15.01.1807, fl. 29 v.

Idioma e escrita

Português

Data de publicação

21/04/2021 00:38:35

id	Date	Attribute	Value	obs
163686	1798-10-15	faculdade	Teologia	Faculdade corrigida
163686	1798-10-15	faculdade-original	Matemática	
163686	1798-10-15	matricula-faculdade	Filosofia	15.10.1798
163686	1798-10-15	matricula-faculdade	Matemática	(obrigado)
163686	1798-10-15	naturalidade	Rio Bom	
163686	1798-10-15	nome	José Doutél	
163686	1798-10-15	nome-apelido	Doutél	
163686	1798-10-15	nome-geografico	Rio Bom	
163686	1798-10-15	nome-nota	frei monge de São Bernardo	
163686	1798-10-15	nome-pai	António Venceslau Doutel	
163686	1798-10-15	nome-primeiro	José	
163686	1798-10-15	ordem-religiosa	Ordem de São Bernardo	frei monge de São Bernardo
163686	1798-10-15	titulo	frei	frei monge de São Bernardo
163686	1798-10-15	titulo	monge	frei monge de São Bernardo
163686	1798-10-15	uc-entrada	1798-10-15	
163686	1799-10-02	matricula-classe	Matemática, 2º ano	02.10.1799
163686	1799-10-02	matricula-faculdade	Matemática	02.10.1799
163686	1799-10-03	matricula-classe	Filosofia, 2º ano	03.10.1799
163686	1799-10-03	matricula-faculdade	Filosofia	03.10.1799
163686	1800-10-03	matricula-classe	Filosofia, 3º ano	03.10.1800
163686	1800-10-03	matricula-faculdade	Filosofia	03.10.1800
163686	1800-10-20	matricula-faculdade	Teologia	(ordinário)
163686	1801-06-11	exame	1º ano	Aprovado Nemine Discrepante, Atos n.º 3, fl. 5
163686	1801-10-23	matricula-classe	Teologia, 2º ano	23.10.1801
163686	1801-10-23	matricula-faculdade	Teologia	23.10.1801
163686	1802-06-18	exame	2º	Aprovado Nemine Discrepante, Atos n.º 3, fl. 34
163686	1802-10-09	matricula-classe	Filosofia, 3º ano	09.10.1802
163686	1802-10-09	matricula-faculdade	Filosofia	09.10.1802
163686	1802-10-15	matricula-classe	Teologia, 3º ano	15.10.1802
163686	1802-10-15	matricula-faculdade	Teologia	15.10.1802
163686	1803-06-08	exame	3.º exame	Aprovado Nemine Discrepante, Atos e Grau n.º 4, fl. 55
163686	1803-10-03	matricula-classe	Teologia, 4º ano	03.10.1803
163686	1803-10-03	matricula-faculdade	Teologia	03.10.1803

id	Date	Attribute	Value	obs
163686	1803-10-22	matricula-classe	Filosofia, 3º ano	22.10.1803
163686	1803-10-22	matricula-faculdade	Filosofia	22.10.1803
163686	1804-06-13	exame	4º exame e Grau de Bacharel	Aprovado Nemine Discrepante, Atos e Grau n.º 4, fl. 96
163686	1804-06-13	grau	Bacharel em Teologia	4º exame e Grau de Bacharel: 13.06.1804, Aprovado Nemine Discrepante, Atos e Grau n.º 4, fl. 96
163686	1804-10-06	matricula-classe	Teologia, 5º ano	06.10.1804
163686	1804-10-06	matricula-faculdade	Teologia	06.10.1804
163686	1805-06-06	exame	5º exame	Aprovado Nemine Discrepante, Atos n.º 4, fl. 135
163686	1806-07-18	exame	Exame Privado e Grau de Licenciado:	Atos Grandes n.º 3, fl. 26
163686	1806-07-18	grau	Licenciado em Teologia	Exame Privado e Grau de Licenciado: 18.07.1806, Atos Grandes n.º 3, fl. 26
163686	1807-01-15	exame	Doutoramento	fl. 29 v
163686	1807-01-15	grau	Doutor em Teologia	Doutoramento: 15.01.1807, fl. 29 v.
163686	1807-01-15	uc-saida	1807-01-15	
163686	2021-04-21	código-de-referência	"PT/AUC/ELU/UC-AUC/B/001-001/D/001409"	
163686	2021-04-21	data-do-registo	2021-04-21	
163686	2021-04-21	url	"https://pesquisa.auc.uc.pt/details?id=163686"	
163686	2022-05-01	nota-processamento	Aviso: faculdade corrigida	Matemática para Teologia.