More time needed

automation, MARC 648, and discovery by time period of subject

James T. Clark 0000-0001-9936-4661
Library Systems and Discovery Manager, Southampton Solent University

Getaneh Alemu 0000-0003-2424-1725
Cataloguing and Metadata Librarian, Southampton Solent University

Received: 12 May 2026 | Published: 16 September 2026

ABSTRACT

This paper describes an investigation into how existing bibliographic metadata can be used to improve discovery by time period of subject. We describe the benefits of using an exact-year date range for the subject matter of the resource, as recommended by FAST (its Chronological heading) and recorded in the MARC 648 field. We assess the feasibility of automatically generating this date range using either AI or conventional scripting techniques. Finally, we describe ways this metadata can be displayed as a facet in discovery systems.

KEYWORDS metadata automation; retrospective cataloguing; generative AI; faceted discovery; time period; MARC 648

CONTACT James T. Clark james.clark@solent.ac.uk Southampton Solent University

1. Facets and time period of subject

To establish the rationale for our investigation, we begin with a description of the current theory and practice in relation to facets in general and specifically faceting by time period of subject (TPS).

1.1. The faceted approach to information organisation

Broughton (2006) notes the difficulty of defining "facet" concisely, but a facet can be understood as a property or characteristic of an information resource (e.g., topic, place, time, form) that enables users to filter, refine, and navigate search results. The origin of the faceted approach in libraries is often identified as Ranganathan's facet analysis in the 1930s, later formalised in his Colon Classification system, which introduced five fundamental categories: Personality, Matter, Energy, Space, and Time (PMEST) (Ranganathan, 1960, 1.25).

Ranganathan's principles were advanced by the UK Classification Research Group (CRG) in the 1950s, influencing systems such as Bliss Bibliographic Classification (BC2)

(Broughton, 2023, p.414). In 1955, the CRG published a manifesto to make faceted classification the "basis for all methods of information retrieval" (Classification Research Group, 1955). And indeed, the faceted approach to classification has also contributed to the development of subject headings and faceted thesauri (Broughton, 2023, p. 411).

1.2. Facets in library discovery systems

Although the strict faceted approaches to information organisation described above have not been widely adopted, the principle of organising subjects into discrete facets has become foundational for modern information retrieval, common in both library discovery systems and commercial platforms (Broughton, 2006, pp. 61–65). Facets' presence in commercial systems presumably helped make them comprehensible to users when they started to appear in 'next generation library catalogues' in the 2000s. Several studies from this period describe facets' popularity with users (Olson, 2007, p. 555; Sadeh, 2008, p. 22; Denton and Coysh, 2011, p. 316). While facets can be a valuable tool in discovery, their value is sometimes severely limited by practical challenges in their implementation that include "user interface design, metadata quality and complexity, the diversity of library resources, and the theoretical and practical impossibility of completely populating consistent, accurate facets" (McGrath, 2023, p. 440).

1.3. Time Period of Subject (TPS) in the library models

In faceted analysis and classification, including Ranganathan's PMEST, Time is one of the fundamental facets and denotes the temporal extent inherent in subject matter rather than a simple record of bibliographic dates. This demonstrates that time period is an important and analytically significant component of subject representation (Dean, 2004, p. 340). The treatment of temporal information in major bibliographic models and metadata standards reinforces this view, consistently recognising time period as a canonical facet.

The IFLA Library Reference Model (LRM) is a conceptual framework that consolidates three earlier models (FRBR, FRAD, and FRSA) to support user tasks such as finding, identifying, selecting, obtaining, and exploring resources. The LRM defines a set of 11 core entities that structure bibliographic and authority data. One of these entities is Timespan, "a temporal extent having a beginning, an end and a duration" (Riva, Le Boeuf, & Žumer, 2017, p. 36). While this entity can apply to many aspects of bibliographic metadata (e.g., date of publication, lifespan of a person) as well as the TPS, its inclusion as a separate entity in LRM emphasises its distinctness as a concept and thus its suitability for use as a facet.

Mirroring LRM, RDA uses the Timespan element to record the beginning, end and duration of temporal extents, including the chronological scope of subject content, while BIBFRAME represents temporal specificity through its *temporalCoverage* property, enabling machine actionable encoding of a work's chronological domain, and Dublin Core provides an equivalent mechanism through its dc:date element, which also supports structured temporal specification. This demonstrates that time period is analytically central to bibliographic metadata and must be recognised as a fundamental entity and facet, not only in theoretical models but also in practical library metadata that supports indexing, filtering and browsing. This position is central to the argument developed in this paper.

1.4. Time Period of Subject (TPS) in discovery

That the TPS of a resource might be of interest to users is fundamental both to the theoretical models for subject analysis and resource description discussed above and also to practical approaches: DDC, LCC, and LCSH all allow for the analysis of material by TPS. Librarians seem to agree: a survey by Hall (2016, p. 113) found that only one out of 40 library staff considered a TPS facet to be less than "moderately useful". On this metric, TPS performed better than other potential facets, including "Person (Subject)", "Geographic Location (Subject)", and "LC Call Number". The importance of TPS for discovery seems a reasonable assumption, but we note that it has not been demonstrated empirically.

The idea of TPS as a facet is not entirely novel. Hall (2011) found that 25.6% of the US university libraries she surveyed had such a facet, while Nahotko (2022, p. 91) found that 21.6% of Polish academic libraries had one. We found these figures surprisingly high, and they are not reflected in the current UK context. In brief searches of the library catalogues of all 24 Russell Group universities, we were unable to find one that had a dedicated TPS facet, though several displayed time periods within a general subject facet.

Some libraries have reported a negative experience with the TPS facet, which might account for its relatively low use. Denton and Coysh (2011, p. 317) report that the TPS facet ("Era") was both sparsely populated and confused by users with the date of publication. Snyder (2010, p. 69), whose research focused on the discovery of music resources, reports that users found the "Time Period" facet to be problematic, as "their labels did not clearly indicate that they contained subject information and ... their content did not seem logical or intuitive, nor was it consistently applicable to music materials." These case studies illustrate some of the challenges with faceting, but they are not fatal to the *principle* of faceting by TPS, as its success often depends on the details of implementation (see 1.2. above). As discovery systems have grown more sophisticated, it seemed a good time to revisit this potentially under-used facet.

1.5. Time Period of Subject (TPS) in LCSH

LCSH is a sophisticated subject vocabulary that has facilitated access to resources across the anglophone world for over 100 years. However, its structure was developed for use in card catalogues, and it is not optimised for modern discovery systems, especially as regards faceting. McGrath (2007) notes several problems that LCSH would cause for a TPS facet; we summarise these below with her examples:

1. Time period is omitted from some subjects where it is considered redundant, e.g., "Underground Railroad$xHistory."
2. Time period is sometimes mentioned or implied in words, e.g., "Iron Age", "Post-Communism", "Literature, Ancient".
3. Time period can often only be entered in pre-selected ranges, not reflecting the actual range covered by the resource, e.g., a book on 1868 might have the subdivision "19th Century."

The difficulties this causes for faceting by TPS are obvious. Resources on a historical topic might have no recorded chronological information to extract into a facet (problem 1); the information might be there but vague and difficult to extract (problem 2); information that is easy to extract might be grossly imprecise (problem 3).

1.6. Time Period of Subject (TPS) in FAST

OCLC developed FAST (Faceted Application of Subject Terminology) by simplifying and reworking LCSH into at first eight, later nine, heading groups, one of which is the Chronological heading (Danskin et al., 2023, pp. 507–508). There are no prescribed time periods in FAST; rather, the Chronological heading is populated with the exact date range that the resource covers. Chan and O’Neill, in an authoritative FAST textbook, state:

> "Chronological headings represent the time period treated or covered in the document or resource at hand. Chronological headings are assigned to reflect the actual time coverage in the resources and are not necessarily limited to specific periods associated with particular events." (Chan and O’Neill, 2010, p. 99)

Thus, in theory at least, FAST avoids the problems that LCSH has with regard to faceting by TPS. The Chronological heading should be present whenever it is relevant, and it should give an accurate reflection of the period that the resource treats.

1.7. LCSH and FAST comparison

The increased precision and consistency offered by using the exact dates makes the FAST approach superior for modern discovery methods, including faceting.[1] The use of exact dates also has the further benefit of simplicity: there are no overheads for the maintenance of a vocabulary, few requirements for training in its application, and it is

[1] See section 3, below, for a discussion of faceting using exact dates. Although future AI-based search may apply filters dynamically, potentially making traditional visual facets and faceted navigation redundant (see Campbell, Jeffery, and Nowicki, 2025), the precision provided by exact date ranges will likely remain relevant.

internationally applicable.[2] Exact dates also seem more suited for applications beyond the purely bibliographic. Increasingly, discovery is happening outside the library, and books are discovered alongside a range of other resource types. In this environment, an approach that can be easily understood out of context and easily applied by non-specialists has distinct advantages.

Similar arguments to those made above have been adduced for the adoption of FAST generally: its simplicity makes it easier to learn and apply, and it is better suited to the modern discovery environment of faceted navigation and linked data (Danskin et al., 2023, p. 507). Our intention here is not simply to restate these arguments, but rather to argue that in the recording of *chronological* data specifically the advantages of FAST over LCSH are greatest. Though simplified from LCSH, most FAST headings are still part of a controlled vocabulary requiring training to apply and continual maintenance; not so the Chronological heading. And it is only in Chronological headings that FAST offers greater precision than LCSH. Furthermore, it is impossible to extract accurate FAST Chronological headings automatically from LCSH; this is not the case for other FAST headings.

Thus, the approach we have adopted in our own catalogue, and that we would advocate for, is to continue to use LCSH alongside a FAST-style exact date range in the 648 when the work has a historical focus.

1.8. FAST Chronological heading: current practice

While in theory the FAST Chronological heading is ideal metadata for a TPS facet, there are problems with the state of current practice. The first is that the headings are not sufficiently widely applied. A search of the UK's National Bibliographic Knowledgebase (NBK) for the 648 MARC field, in which the heading is recorded, found it in just 0.7% of records. By contrast, the relevant subfields for LCSH chronological subdivisions (650$y and 651$y) are present in 8.3% of records. Furthermore, it is likely that many 648 fields do not contain FAST data. From a sample of 11,879 MARC 648 fields only 4,375 (36.8%) were coded as FAST. Thus, the proportion of FAST Chronological headings in the NBK might be as low as 0.25%.

The second problem is that very often the value in the FAST Chronological heading does not in fact attempt to record exact dates for the TPS but instead has been automatically generated from the dates in the LCSH. Thus, a book exclusively covering the 1970s is given the Chronological heading "Since 1960" in FAST, this being a combination of the earliest and latest dates found in its LCSH ("$y1960-1980" and "$y1969-"). Of the 4,375 FAST Chronological headings in our NBK sample, 470 (10.7%) end "-2099" and 227 (5.2%) are open-ended ("Since ..."). These headings, at least,

[2] In this context, it is worth noting that the *Regeln für die Schlagwortkatalogisierung* (RSWK), the German subject headings system, also tends to assign exact dates in the 648 field (Frommeyer, 2004, p. 202); the National Library of Poland Descriptors (NLD) uses both exact dates (in the 045 field) and preassigned ranges that apply across the whole vocabulary, in the 648 field (Cieloch-Niewiadomska, 2019, p. 52).

appear to have been automatically converted from LCSH, and it is likely that many others have too. This is in line with guidance: the FAST Converter tool provided by OCLC generates such headings[3], and Chan and O'Neill (2010, p. 176) recommend the use of these "when more specific dates are not available." This approach, which allows automatic conversion from LCSH to FAST, is presumably founded on the belief that this data is better than nothing. We agree, but it is unfortunate that the provenance of this data, that it was automatically generated from LCSH, was not recorded.

2. Automatically populating the 648 field

We have seen that an exact date TPS currently exists in only a tiny proportion of records, and manual retrospective addition of the data at that scale is clearly out of the question. Therefore, we decided to investigate the feasibility of fully automating the generation of this metadata, i.e., finding an automation method that is sufficiently accurate (for a subset of records, at least) to be applied without human oversight at the record-by-record level.

2.1. Trying generative AI: methodology

Given the current excitement (or perhaps hype) about the potential of generative AI (GenAI) to 'understand' natural language, we decided to investigate whether GenAI tools could accurately create an exact TPS based on existing metadata. We chose Google's Gemini as our tool, as its API allowed generous usage limits without charge.

2.1.1. Generate a set of sample records

After early experiments with GenAI, we decided to provide three MARC fields to the tool: the title (245, minus $c), contents (505), and summary (520). We chose not to provide subject headings as the date ranges there can be imprecise. In selecting a sample of records to work with, we merely wanted records that were likely to have historical subjects and that had some indication of a date in these fields. We located these by a crude methodology, first searching our catalogue for the keyword "history", then filtering for 'date terms', i.e., four consecutive numbers (i.e., a year) or the string "centur" (i.e., the word "century" or "centuries"). This approach returns some false positives, but this was not a concern. We also needed records to have a single publication date (extracted from the MARC 008) to use as the upper bound.

As a final criterion, we excluded two types of material that early experiments suggested the tool would perform badly on: multi-volume works and exhibition catalogues. The problem with multi-volume works is that the dates mentioned in our selected metadata fields often refer to the period covered by the whole work, while we might only own a single volume. Often the only date that records for exhibition catalogues contain is the date of the exhibition, and arguably this is not relevant to the TPS.

[3] Available online at https://fast.oclc.org/lcsh2fast/.

Our initial search for the keyword "history" retrieved 10,955 records. The further filters reduced the sample to 4,491 records. We treated these as three separate sets: those with a 'date term' in the 245 only (516 records), those with a 'date term' in the 245 and at least one of the 505 and 520 (974 records), and those with no 'date term' in the 245 but in at least one of the 505 and 520 (3,001 records).

2.1.2. Push through Gemini with prompt

These records were pushed through the Gemini API with the following prompt, refined from early experiments:

> "Below is the title, table of contents, and summary of a student essay. Give me a date range in years for the subject-matter of the essay and a percentage value for how confident you are that this date range is correct. For example, if there was nothing in the data to indicate the date range, you could return a confidence level of 0%. Give me the date range only first, then confidence percentage only, then any further information. Separate these three sections with the tab character."

The first thing to address is our referring to the resource as a "student essay". We were concerned that if we described the resource as a book, the tool would search for information about it, find bibliographic data, and return the dates from its LCSH. We did not run a controlled experiment to find if our approach was necessary or had any impact, but anecdotally we noticed no evidence of LCSH dates being returned by the tool.

The prompt displays some features of a good prompt: it gives some context and, while the text could be sharper and more concise, it has no ambiguity and specifies the output format. However, we did not use advanced techniques of prompt engineering such as scenario framing ("imagine you are a library cataloguer") or temperature setting. The great number of different patterns to be found in our input data made us sceptical of the value of providing a worked example ("shot") or chain-of-thought reasoning, and our early experiments did not change our mind.[4] But no doubt the prompt could be improved in some respects.

2.1.3. Process and filter the results

We anticipated that the tool would not be able to produce accurate data in every case, and we were keen to find ways to automatically identify data that was more likely to be accurate. One way was to ask the tool to rate its confidence in the result. We also decided to run the queries through the tool twice and see whether the same result was obtained. We left approximately a month between the two query runs.

[4] See Khan, 2024, pp. 35–40 for the characteristics of a good basic prompt and (p. 50) on temperature setting. See Debnath et al., 2025, pp. 6–8 for the concept of "shots" and chain-of-thought prompting.

Once we had results from the two runs, several short Python scripts were used to process the results. First, the few occasions where the tool had failed to return a valid result (either a single year or a year range) were excluded. The second process checked the upper bound (if one existed) from the returned date range against the publication date, choosing the earlier of these two dates as the new upper bound. Finally, we checked the results from the two different runs against each other; only the results where the two runs matched were passed to the validation stage.

2.1.4. Validate for acceptability

Random samples were taken from the result sets for validation. Our validation method was to check the date ranges for acceptability, that is, to check that they did not misrepresent the resource. This differs from the gold-standard approach to validating GenAI tools, which would have meant assigning dates to the resources independently of the tool and then checking for matches. The gold-standard approach is a stricter test, but we do not think it is a more appropriate one here.[5] What we are interested in, after all, is how often the tool can apply useful metadata, not how often it can imitate a human. Furthermore, checking for acceptability acknowledges that there is a subjective element in assigning a date range to the subject, and it is the approach we would take if we were assessing the work of a new human cataloguer.[6]

The gold-standard approach to validation would have been superior in one respect: it removes the potential for validators to skew the results through some (possibly unconscious) desire for the tool to appear either effective or ineffective. We tried hard to be objective in our judgements of acceptability, but it is impossible for us to say whether we were successful; the best test of this is whether our results are replicable.

Unless there was a strong reason not to, we considered self-defined date ranges – those stated explicitly in the title or summary – as acceptable. And when the metadata we had for the resource was very full, we did not always consult the full text. The logic here was that even if we had the book in hand, we would form our judgements initially on the title, the contents, and the summary – the very metadata that we had – and would not necessarily scan the main text if we found those elements informative enough. In one respect we took a consciously lenient approach: if the upper bound was the publication date of the resource, we deemed this acceptable even if the upper bound should have been a year earlier. This is because it would probably have been more sensible to apply the publication year minus one as an upper bound, given the usual gap between a manuscript being closed and the book being published.

[5] Others investigating the use of AI for subject analysis also deviate from the gold-standard approach: Chow, Kao, & Li (2024, p. 582) use an acceptability validation, while Dobreski and Hastings (2024, p. 5) check against a gold standard but also pass acceptable headings.

[6] Deciding on an acceptable date range is not a simple task. Consider a book of 300 pages that mainly covers the years 1830-1850, but whose ten-page introduction reviews the years 1780-1830. Should we accept the date range 1780-1850 for this book? Probably not, but what if the introduction is 60 pages and thus 20% of the resource? And then what if the years 1780-1820 are dealt with in the first 40 pages of those 60?

2.2. Trying generative AI: results

Table 1 gives the results of our investigation. The three rows represent the three input sets obtained from the process described above (2.1.1.). The first column gives the number of records in the set. The second column gives the number of records that progressed to validation, that is, for which we got a valid and matching result from both runs through the AI tool. So, in the case of those records with a date term in the 245 only, a valid matching result was produced for 449 of the 516 records (87%).

The next column ("Sample 1 acceptability") shows the acceptability rate for the first sample we took, which for this group was 97%. We noticed that when the combined confidence level the tool produced for the result was 198 (i.e., 99% on both runs, or 98% on one and 100% on the other) or above, the results were more accurate, so we took a second sample that included only these results. In validating this set we noticed that when the result was a single date, rather than a date range, it was often faulty, so we excluded these too. The next column ("Sample 2 acceptability") shows the rate of acceptability here. The final column shows the number of results that meet the criteria for sample 2, i.e., that were available for sampling.

Group	Number of records	Two AI answer match	Sample 1 acceptability	Sample 2 acceptability	Meet criteria for Sample 2
Date in 245 only	516	449 (87%)	97/100 (97%)	197/199 (99%)	350 (68%)
Date in 245 and 505/520	974	821 (84%)	95/100 (95%)	187/195 (96%)	628 (64%)
Date in 505/520	3,001	1,654 (55%)	68/100 (68%)	86/100 (86%)	394 (13%)

Table 1: Acceptability rates of AI-generated date ranges

The groups with a date in the 245 (the top two rows in the table) performed very similarly. The tool’s responses from both runs matched in around 85% of cases, and the acceptability rate for Sample 1 was 95% and above. We can increase the acceptability rate a little by applying the filters that produced our Sample 2, but there are diminishing returns here: the acceptability creeps up a few percentage points, but the proportion of records that now meet the criteria drops to around 65% (the final column in the table).

Records without a date in the 245 (the bottom row) perform less well. We can reach an acceptability rate of 86% for Sample 2, but then only 13% of records meet the criteria.

2.3. Trying generative AI: errors

Table 2 gives examples of errors that persisted into our Sample 2. Each one illustrates a different kind of error. For *An economic history of Europe*, the tool has been unable to recognise from "and the subsequent crisis" that the book treats the period beyond 2008. For *Pocketbook politics* it incorrectly applies the date 1970 as the end of the "Great Inflation of the 1970s." Even if it had done better here, we would still have considered the date range unacceptable, as the book mainly deals with 1900-1960, the period after 1960 being confined to a short epilogue. *Style 1930* treats only 1930 and the few years either side, not the whole 1930s. The real error here is one of over-confidence; with such limited information we would expect the confidence rating to be lower than the threshold required for our sample.

Title	Extract of 505/520	Gemini response
An economic history of Europe since 1700	… The book ends with an assessment of the impact of the global crash of 2008 and the subsequent crisis of the Eurozone …	1700-2008
Pocketbook politics : economic citizenship in twentieth-century America	… ending with the Great Inflation of the 1970s.	1900-1970
Style 1930		1930-1939

Table 2: Representative errors in AI-generated date ranges

It is notable that we have seen instances where the tool dealt well with situations just like these, so there is a certain amount of apparent randomness to the responses. The tool performs much more inconsistently than a human would. The quality of input data is important, but even apparently good quality input data can be misleading, as in the Pocketbook politics example. In combination, these two factors mean that whether the tool gives an acceptable response involves a fair amount of chance. This might be disconcerting, but it does not necessarily obviate its use: the key is to find the right balance of filters that means the tool's failures occur at an acceptably low rate. A further important point is that in most cases when the response was deemed unacceptable, it was wrong by only a few years and, we suspect, closer to the truth than LCSH time periods would have been.

2.4. Trying regular expressions

The AI tool's results for those records that have a date in the 245 may seem impressive, but of course many of these records include an explicit declaration of the date range, such as *Women's suffrage in Britain 1867-1928*. The tool is not doing anything especially sophisticated in extracting this range, and it is possible to do the same using regular expressions. We used a Python script to identify a date range from 14 different regular expression patterns, such as "in the … century" (*The British working* *class in the twentieth century*), "of the [decade]" (*Living up to the ads : gender fictions of the 1920s*), and "[date] to now" (*Photography at MoMA : 1960 to now*).

This approach generated date ranges for 1,077 of the 1,490 records (72%), and manual checking revealed that in only two instances did this date range fail to reflect the title of the resource, due to unanticipated patterns in the data. There is scope for this approach to be improved still further by accounting for more complex patterns and perhaps incorporating natural language processing (e.g., to identify parts of speech).

2.5. Review

When records include a date in the title, automatically generating a TPS is relatively simple. Regular expressions seem to give the best results, with an acceptability of almost 100% for 72% of records, though GenAI is not far behind. Once the low hanging fruit that can be handled by regular expressions is removed, GenAI is less successful, but it can achieve an acceptability rate of 91% (97/107) for 26% of the records with a date in the title that our regular expression script passed over, and 86% (86/100) for 13% of records with no date in the title.

3. Discovery

3.1. Faceting with raw data: problem and solution

There is an obvious difficulty in presenting exact date ranges in a facet: the user will see many overlapping date ranges and need to select multiple options. The predefined time periods in LCSH alleviate this a little, but with these too there will be considerable overlap; works about the 1500s might, depending on their main subjects, have time periods defined as "16th century", "Early modern and Elizabethan, 1500-1600", "Early modern, 1500-1700", "Renaissance, 1450-1600", "Wars of the Huguenots, 1562-1598", etc.

The solution is similarly obvious: to allow the user to type in a date range (or use a slider to select it). Such tools are common on commercial websites, normally for selecting a price range, and already exist on some library discovery systems for filtering by publication date. These tools will need adapting to work with a date range (as opposed to a specific date or dates), but this should be well within the capabilities of library software providers. Such a tool could also be used for other chronological facets, such as the creation dates now recorded in the 046 MARC field (Mullin, 2024, pp. 138–139).

3.2. Faceting with processed data: an interim solution

But we do not have to wait for the perfect faceting tool before using this data. At Solent we developed an interim solution that takes the year range from the 648 field and converts it to two groups of data, stored in local fields in the MARC record: one group contains the centuries covered by the year range, the other contains the decades covered by it. Thus a 648 of 1756-1779 would be given "18th century" in the centuries field and "1750s", "1760s", and "1770s" in the decades field. This conversion work is fully automated via normalisation rules and the Alma API. As in most instances LCSH time periods identify centuries correctly, the script also generates centuries (but not decades) for records that have LCSH time periods but no 648.

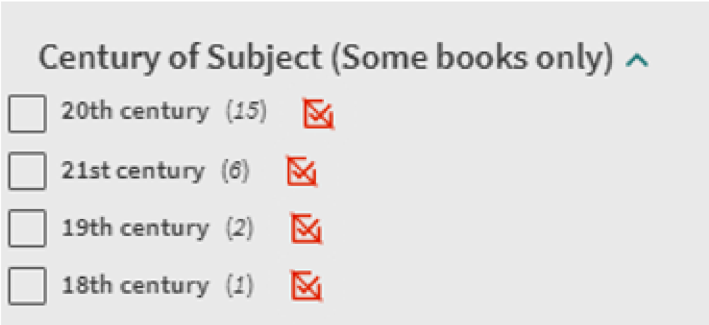

Figure 1: Century of subject facet in Solent's catalogue

Once created, the century and decade fields are then used as facets in the catalogue (see Figure 1). The level of precision used here is a matter of choice; with a small tweak the script could output periods of, say, 25 or 50 years, instead of decades and centuries. As discussed, this is not our ideal solution, and there is always some maintenance overhead to a customisation, even when fully automated like this one. But this shows what is possible with existing technology. These facets went live in Solent's catalogue in February 2026, and we will monitor their use. A prototype of the centuries facet was available for the previous nine months. In absolute terms it was not heavily used, but once allowance is made for the fact that it appears as a facet on only about 20% of searches (due to the small proportion of records where TPS is relevant), it was used about half as frequently as the general subject term facet. This is not a bad return, considering time period is only one aspect of a subject, and it justifies its continued inclusion in our catalogue.

4. Conclusions

4.1. Reflections on Time Period of Subject (TPS) data

Describing the TPS of a resource is a fundamental part of subject analysis. We are convinced that in the contemporary discovery context, the best way to do this is through assigning the exact date range of the subject matter.

The main obstacle to adopting this practice is the enormous number of records that are missing this data. To combat this, FAST allows the use of LCSH time periods as an alternative, allowing automated conversion from LCSH to FAST. Despite the problems with LCSH time periods, we approve of this approach on the basis that this data is better than nothing. But the publication date of the resource should be used as the closing date of the time period, thus avoiding date ranges such as "2000-2099". And more importantly, the provenance of the data should be identified, e.g.,

> 648 #7 $a1960-2000$2fast$7(dpsfa)Derived from LCSH subdivisions and publication date.

Identifying the provenance of AI-generated data is widely acknowledged to be important, and it is similarly important for other automated methods that are prone to producing errors. Recording data provenance will make it easier to revisit the LCSH-derived dates later and improve upon them, either manually or by some automated process.

In this project we investigated the potential for automatically generating exact time periods using three key metadata fields: the title, contents, and summary. We were able to generate this data with a high degree of confidence for about a quarter of records. We note, however, that most successes came from the 'low-hanging fruit', where the date was mentioned in the title; when this information is absent, the success rate drops dramatically.

The project also investigated methods for the use of this data in modern discovery tools. While a 'date slider' facet appears best suited for use with this data and is presumably not beyond the capabilities of library systems providers, we were able to demonstrate that a workable alternative is available now.

4.2 General reflections on automated metadata generation

It is a commonplace to insist that GenAI in cataloguing must be used with human supervision.[7] We agree, but we contend that the kind of supervision used here, where quality is maintained by sampling, can be sufficient.[8] This project looked at the retrospective addition of the 648 field to older MARC records. In the vast majority of records, this field has not previously been added by human cataloguers, and there is no realistic prospect of the cataloguing community achieving this retrospective addition manually. In cases like this, then, the question is whether we would rather have no data at all or data that is potentially erroneous. That will depend on factors such as the potential harm done by erroneous metadata, the value added by the extra metadata, and the proportion of errors. In this particular case, where we can generate the 648 for a subset of records with an acceptability rate above 85%, we think the automated approach deserves further investigation. It might put this question in some sort of perspective to remember that the guidance for the FAST Chronological heading is that LCSH time periods should be used where an exact date range is not available (Chan and O’Neill, 2010, p. 176). The date ranges generated by this approach are almost always considerably less satisfactory than the 'unacceptable' ranges generated by our AI tool. If the imprecise ranges derived from LCSH are deemed better than

[7] For bibliography, see the review article by Engel et al., 2025, who conclude (p. 8), "there is general agreement that human oversight is a necessary component for the use of AI technologies today."

[8] The German National Library have used this approach for some time; see Mödden, 2022, p. 262.

nothing, then a smaller proportion of more minor errors from GenAI should not cause more concern.

Many investigations into cataloguing applications of GenAI look for ways to make what we already do more efficient. This is important work, of course, but we should also look for ways that GenAI can allow us to do more, to add new or rarely used metadata elements. That is what this project does, in seeking to generate retrospectively a field that exists in only a tiny proportion of records currently. There are historical instances of librarians responding to technological change by expanding their scope (Abbott, 1988, pp. 138–184 and pp. 138–120), and GenAI provides such an opportunity to cataloguers.

We have elsewhere stressed the value of the 505 and 520 fields in MARC records for helping users find and select resources (Alemu, 2022, pp. 183–14; Clark, 2024, p. 196). This project has shown that they are also useful in the automatic generation of further metadata. In the absence of full text (and possibly even when it is available), ensuring that good quality content and summary fields are available for as many records as possible should be foundational to any attempt to automate subject analysis.

Finally, while we are generally positive about the potential of GenAI both in this particular application and in general, we would temper that with an acknowledgement of the need for caution.[9] In particular, we point to the way that simple regular expressions returned better results for records that had a date in the title. If we intend to embrace technology, then we should embrace a suite of tools, and not rush headlong for GenAI every time.

References

Abbott, A. (1988) *The system of professions: An essay on the division of expert labor*. University of Chicago Press.

Alemu, G. (2022) *The future of enriched, linked, open and filtered metadata: Making sense of IFLA LRM, RDA, Linked Data and BIBFRAME*. Facet Publishing.

Broughton, V. (2006) 'The need for a faceted classification as the basis of all methods of information retrieval', *Aslib Proceedings*, 58(1–2), pp. 49–72. Available at: https://doi.org/10.1108/00012530610648671 [Accessed: 10 August 2026]

Broughton, V. (2023). 'Facet analysis: The evolution of an idea', *Cataloging and Classification Quarterly*, 61(5–6), pp. 411–438. Available at: https://doi.org/10.1080/01639374.2023.2196291 [Accessed: 10 August 2026]

Campbell, L., Jeffery, K., and Nowicki, R. (2025) 'AI, acronyms, and the future of facets: Mnemonics as heuristic instructional tools for structuring GenAI prompts', *Qualitative and Quantitative Methods in Libraries*, 14(1), pp. 121–134. Available at: https://www.qqml-journal.net/index.php/qqml/article/view/906 [Accessed: 10 August 2026]

[9] For a balanced assessment of the use of GenAI in cataloguing, see Moulaison-Sandy and Coble, 2024.

Chan, L. M., and O'Neill, E. T. (2010) *FAST: Faceted application of subject terminology: Principles and applications*. Libraries Unlimited.

Chow, E. H. C., Kao, T. J., and Li, X. (2024) 'An experiment with the use of ChatGPT for LCSH subject assignment on electronic theses and dissertations', *Cataloging and Classification Quarterly*, 62(5), pp. 574–588. Available at: https://doi.org/10.1080/01639374.2024.2394516 [Accessed: 10 August 2026]

Cieloch-Niewiadomska, J. (2019) 'Introducing the National Library of Poland descriptors to the Polish national bibliography', *Cataloging and Classification Quarterly*, 57(1), pp. 37–58. Available at: https://doi.org/10.1080/01639374.2019.1573774 [Accessed: 10 August 2026]

Clark, J. T. (2024) 'The format and display of the MARC 505 contents note: Some recent practical developments', *Cataloging and Classification Quarterly*, 62(2), pp. 188–208. Available at: https://doi.org/10.1080/01639374.2024.2350467 [Accessed: 10 August 2026]

Classification Research Group (1955) 'The need for a faceted classification as the basis for all methods of information retrieval', *Library Association Record*, 57(7), pp. 262–268.

Danskin, A., et al. (2023) 'FAST the inside track: Where we are, where do we want to be, and how do we get there?', *Cataloging and Classification Quarterly*, 61(5–6), pp. 506–524. Available at: https://doi.org/10.1080/01639374.2023.2215762 [Accessed: 10 August 2026]

Dean, R. J. (2004) 'FAST: Development of simplified headings for metadata', *Cataloging and Classification Quarterly*, 39(1–2), pp. 331–352. Available at: https://doi.org/10.1300/J104v39n01_03 [Accessed: 10 August 2026]

Debnath, T., et al. (2025) 'A comprehensive survey of prompt engineering techniques in large language models', *TechRxiv* [Preprint]. Available at: https://doi.org/10.36227/techrxiv.174140719.96375390/v2 [Accessed: 10 August 2026]

Denton, W., and Coysh, S. J. (2011) 'Usability testing of VuFind at an academic library', *Library Hi Tech*, 29(2), pp. 301–319. Available at: https://doi.org/10.1108/07378831111138189 [Accessed: 10 August 2026]

Dobreski, B., and Hastings, C. (2025) 'AI chatbots and subject cataloging: A performance test', *Library Resources & Technical Services*, 69(2). Available at: https://doi.org/10.5860/lrts.69n2.8440 [Accessed: 10 August 2026]

Engel, J. Y., et al. (2025) 'Artificial intelligence in library cataloging: A review of literature', *Journal of Library Metadata*, 25(4), pp. 261–276. Available at: https://doi.org/10.1080/19386389.2025.2526913 [Accessed: 10 August 2026]

Frommeyer, J. (2004) 'Chronological terms and period subdivisions in LCSH, RAMEAU, and RSWK', *Library Resources & Technical Services*, 48(3), pp. 199–212. Available at: https://doi.org/10.5860/lrts.48n3.199-212 [Accessed: 10 August 2026]

Hall, C. E. (2011) 'Facet-based library catalogs: A survey of the landscape', *Proceedings of the American Society for Information Science and Technology*, 48(1). Available at: https://doi.org/10.1002/meet.2011.14504801195 [Accessed: 10 August 2026]

Hall, C. E. (2016) *Facets in library catalogs: The beliefs, behaviors, policies and practices that guide implementation*. PhD thesis. Drexel University. Available at: https://doi.org/10.17918/etd-7078 [Accessed: 10 August 2026]

Khan, I. (2024) *The quick guide to prompt engineering*. Wiley.

McGrath, K. (2007) 'Facet-based search and navigation with LCSH: Problems and opportunities', *Code4Lib Journal*, 1. Available at: https://journal.code4lib.org/articles/23 [Accessed: 10 August 2026]

McGrath, K. (2023) 'Musings on faceted search, metadata, and library discovery interfaces', *Cataloging and Classification Quarterly*, 61(5–6), pp. 439–490. Available at: https://doi.org/10.1080/01639374.2023.2222120 [Accessed: 10 August 2026]

Mödden, E. (2022) 'Artificial intelligence, machine learning and bibliographic control: DDC short numbers – Towards machine-based classifying', *JLIS.it*, 13(1), pp. 256–264. Available at: https://doi.org/10.4403/jlis.it-12775 [Accessed: 10 August 2026]

Moulaison-Sandy, H., and Coble, Z. (2024) 'Leveraging AI in cataloging: What works, and why?', *Technical Services Quarterly*, 41(4), pp. 375–383. Available at: https://doi.org/10.1080/07317131.2024.2394912 [Accessed: 10 August 2026]

Mullin, C. A. (2024) *Many pathways for discovery: Describing music resources using faceted vocabularies*. Music Library Association & A-R Editions.

Nahotko, M. (2022) 'Knowledge organization affordances in a faceted online public access catalog (OPAC)', *Cataloging and Classification Quarterly*, 60(1), pp. 86–111. Available at: https://doi.org/10.1080/01639374.2021.2015734 [Accessed: 10 August 2026]

Olson, T. A. (2007) 'Utility of a faceted catalog for scholarly research', *Library Hi Tech*, 25(4), pp. 550–561. Available at: https://doi.org/10.1108/07378830710840509 [Accessed: 10 August 2026]

O'Neill, E. T., and Chan, L. M. (2003) 'FAST (Faceted application of subject terminology): A simplified LCSH-based vocabulary', *World Library and Information Congress: 69th IFLA General Conference and Council*, Berlin. 219–2 August. International Federation of Library Associations and Institutions. Available at: https://webdoc.sub.gwdg.de/ebook/aw/2003/ifla/vortraege/iv/ifla69/papers/010e-ONeill_Mai-Chan.pdf [Accessed: 10 August 2026]

Ranganathan, S. R. (1960) *Colon classification*. 6th edn. Asia Publishing House.

Riva, P., Le Boeuf, P., and Žumer, M. (2017) *IFLA Library Reference Model: A conceptual model for bibliographic information*. IFLA. Available at: https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017.pdf [Accessed: 10 August 2026]

Sadeh, T. (2008) 'User experience in the library: A case study', *New Library World*, 109(1–2), pp. 7–24. Available at: https://doi.org/10.1108/03074800810845976 [Accessed: 10 August 2026]

Snyder, T. (2010) 'Music materials in a faceted catalog: Interviews with faculty and graduate students', *Music Reference Services Quarterly*, 13(3), pp. 66–95. Available at: https://doi.org/10.1080/10588167.2010.528746 [Accessed: 10 August 2026]