

Death and Taxonomies: Bringing Certainty to Legal Language

Abstract: Taxonomies are key to creating certainty in language across the legal landscape. In this article **Alice Laird** and **Katy Snell** of Howard Kennedy examine why taxonomies are still important in the age of AI. They also discuss how they have built on their expertise on taxonomies for the benefit of their firm, and how they have become involved with a cross firm project called noslegal.

Keywords: taxonomies; information retrieval; artificial intelligence; AI

Benjamin Franklin once said, "in this world nothing can be said to be certain, except death and taxes".¹ In this article we would like to present an addition to that list. Without getting too philosophical, one of the ways we encourage certainty is through creating consistency and thus we offer you taxonomy as a finishing flourish to that idiom. In an age where LLMs (large language models) and ChatGPT seem to be the only topic at a legal tech networking event, the need for this will become more important than ever. There is no doubt that the technology is impressive and has the potential to make changes to the way we work; however, at the end of the day the information coming out is only as good as the data being searched or interrogated.

Taxonomies are a way of bringing order to legal language. In the text below we will talk in more detail about why having a taxonomy is so important, the Howard Kennedy journey so far and, if you haven't heard about it already, a brilliant cross law firm project called noslegal, which we have already started using (<https://www.noslegal.org/>).

WHAT IS A TAXONOMY AND WHY DO WE NEED THEM?

This isn't an article about generative AI – there's certainly enough of those already – but at the point of writing this, every alert, conference speech and LinkedIn feed were alight with interest in the topic. It would be remiss not to at least mention how the two things fit together. There is certainly a buzz that AI is going to be the answer to search problems and eliminate all the dreary elements of our jobs, but we caution you to think again. One of the main issues coming up in ChatGPT is the concept of hallucinations – the fact that it can generate seemingly plausible ideas or content that are completely made up. By now everyone will have seen the case in the US where a lawyer quoted cases they found via the program only to discover that they were completely made up.² Some of the more experienced minds in legal tech have already started looking to employ this kind of technology in structured data sets to bring back better quality results. We believe that this is where some of the greatest successes are going

to come about – where the data being scanned is already organised and labelled in an effective, consistent way. Which brings us to taxonomies. To get a greater understanding of what they actually are it is interesting to look outside law firms to observe and learn from some established and innovative examples of taxonomies.

Taxonomies are everywhere. We have included a look at some of the ways taxonomies are applied to assist with everyday tasks and projects. Some are very simple and long-established; others harness technological innovation to underpin artificial intelligence improvements. We also examine a couple of specific use cases in more detail.

TAXONOMIES IN EVERYDAY LIFE – SUPERMARKETS, SHOPPING WEBSITES, MUSIC

We are all familiar with the role that taxonomies play in adding structure and organisation to our lives. For example, we know that we can find frozen peas in a supermarket in the frozen vegetables section, as we know that peas are classified as a vegetable and that they can be successfully frozen.

We rely on ever-more complex product taxonomies on shopping websites, supported by detailed metadata and tagging, to find the exact item that matches our requirements.

Taxonomies are also used in music publishing metadata to assist consumers in accessing music or podcasts from the selected genres, composers and artists. This metadata is also key in ensuring that payments are made to artists. Bad metadata leads to the well-reported problems with these taxonomies, resulting in errors for consumers and artists – consumers do not get the music they are expecting; artists do not get paid.

MEDICAL AND ACCOUNTANCY TAXONOMIES

We have chosen a couple of examples which show in more detail how taxonomies are used in the medical and



Finding the specific aisle you need in the supermarket through unconscious use of classification schemes

accountancy professions. We read about the SNOMED example and think it is fascinating and has the potential to improve health experiences for whole populations. We selected the ICAEW (Institute of Chartered Accountants in England and Wales) use case because it is a project that Alice worked on and is now integral to how ICAEW's website content is managed and leveraged.

SNOMED CT – USING A TAXONOMY WHEN EXPERTS IN VARIOUS SETTINGS USE DIFFERENT WORDS TO REFER TO THE SAME THING

Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) is described as a structured clinical vocabulary / taxonomy for use in an electronic health record. It is the most comprehensive and precise clinical health terminology product in the world, forming an integral part of the electronic care record. It represents care information in a clear, consistent and comprehensive manner.³

SNOMED CT allocates a concept label and unique ID for terms related to the direct management of care of an individual including, terms related to diagnosis and procedures, symptoms, family history, allergies, assessment tools, observations, devices.

Different healthcare workers and settings, for example dietitians, pharmacists, doctors and nurses, may use a range of terminology to describe a concept. These

variations are all captured as synonyms in SNOMED. Use of any of the synonyms in an individual's care plan will link back to the same concept. Specialists can use terms they are familiar with, and vital information is shared consistently across health and care settings. In England SNOMED CT was implemented within GP practices in a phased approach from April 2018.

SNOMED CT is a multinational and multilingual terminology which can manage different languages and dialects.

ICAEW – UNDERSTANDING WHAT IS HAPPENING BEHIND THE SCENES WHEN A TAXONOMY IS APPLIED TO A WEBSITE OR OTHER KNOWLEDGE SETTING

ICAEW uses an in-house-developed, topic-based corporate taxonomy covering the world of accounting and finance. As content is added to the website, rules-based auto-classification with natural language processing uses APIs (Application Programming Interface) to integrate taxonomy management software to content management software and auto-tags content with terms that describe the subjects of the web page text.

Queries are then used to combine terms from different taxonomy facets to present content on individual webpages – for example, a page powered by a taxonomy query will show all the latest webinars available on



Example filters for a clothing shopping website

levelling-up or the most recent articles on cyber-security. Automated content can be combined with manually selected content. The approach successfully combines content on a subject produced by different departments in the organisation. The taxonomy is also a useful tool in discovering subject areas lacking up-to-date content.

An advantage of the way that ICAEW's taxonomy works is that the content auto-classification and pre-set queries mean that content management systems' (CMS) editors do not need to interpret and manually remember and apply classification every time they add content. However, this process has not removed the need for skilled involvement from information professionals. Building the taxonomy and setting up weighted classification rules was a lengthy project. Once the project stage was complete and the taxonomy integrated, jobs for the team include regular checks on auto-classification and queries – the classification works well but can occasionally throw up anomalies or needs fine-tuning via additional

synonyms. The taxonomy needs constant review as new subject areas develop. Also, as the software develops, classification rules need to be updated to take full advantage of improvements.

TAXONOMIES IN EVERYDAY LIFE: INNOVATIVE TECHNOLOGIES – KNOWLEDGE GRAPHS

Taxonomies are often an integral part of innovative technologies, such as Knowledge Graphs. Knowledge Graphs are excellent tools for the retail, finance, healthcare and entertainment industries as they organise data from multiple sources, often via taxonomies. They capture information about people, places or events and forge connections between them.⁵ Google Knowledge Panels are powered by Knowledge Graphs.

DBpedia is a project that extracts structured content from the information created in the Wikipedia project. This structured information knowledge graph, using taxonomies, is made available on the World Wide Web, and automatically evolves as Wikipedia changes. You can use it to ask queries against Wikipedia, for instance "Give me all cities in New Jersey with more than 10,000 inhabitants," or, "Give me all Italian musicians from the 18th century".

Enterprises, such as Apple (via Siri), Google (via Freebase and Google Knowledge Graph), and IBM (via Watson), and particularly their respective high-visibility projects associated with artificial intelligence, have benefited immensely from DBpedia's contributions.⁶

A taxonomy describes the domain of information through a standardised vocabulary, capturing synonyms and alternative ways of describing the same concept, as well as basic relationships between concepts (broader, narrower, and related terms). This initial structuring of an information domain can then be leveraged by knowledge graphs and AI, making taxonomies an essential building block for advanced semantic solutions. Organisations that understand the value of taxonomies and how they enable future technologies will be better poised to take advantage of AI and semantic tools.⁷

HOWARD KENNEDY'S JOURNEY

Most innovations come about because a critical number of things accumulate at once which mean action can no longer be avoided, and this is exactly how our journey at Howard Kennedy started. Several internal projects occurred in quick succession which highlighted a need

SNOMED CT Taxonomy example

Fully specified name	Concept ID	Defining relationship	Qualifiers	Synonyms
Common cold	82272006	Is a: Virus upper ... Causative agent: Virus ... Finding site: upper res ...	Qualifying characteristics and facts	Common cold (preferred) Acute coryza Acute nasal catarrh

Artificial intelligence

Discover more about the impact of artificial intelligence and the opportunities it presents for the accountancy profession.

In this section

Features and articles

[See all our features and articles >](#)

ICAEW urges clarity in AI regulatory consultation

ICAEW's consultation response supports the UK government's ambition for flexible AI regulation, but warns clearer guidelines for effective implementation and compliance are needed.

ICAEW Insights
ICAEW

Article Viewpoints June 2023
29 June 2023

AI takeover, part 2: ethical and regulatory implications for financial services

Polly Tsang, Financial Services Manager at the ICAEW navigates AI's potential to reshape financial services in a two-part series. In the second part, she questions the ethics and regulation of the use of AI in financial services.

ICAEW Insights
ICAEW

Article Viewpoints June 2023
13 June 2023

AI takeover, part 1: risks and opportunities for financial services

Polly Tsang, Financial Services Manager at the ICAEW navigates AI's potential to reshape financial services over a two-part series. First up: the risks and opportunities to the industry.

ICAEW Insights
ICAEW

Article Viewpoints June 2023
12 June 2023

Tech and the future of M&A

Harnessing the opportunities technology offers M&A is a challenge. David Prosser speaks to advisers and investors about new technology and its impact on the future of dealmaking and execution.

ICAEW
Corporate Financier
ICAEW

Article 09 June 2023

An example of an ICAEW webpage with content surfaced using taxonomy facets

for greater data governance. They all involved fields and data which might need to be used in other systems but there was no internal agreement on what these should be. The idea of having some controls over terms was raised but not much came of it until summer 2022 when the Knowledge Team invited a consultant, CJ Anderson of Iron Carrot,⁸ to write a report on what benefits a formal taxonomy could provide for the firm.

Interviews carried out with staff in key stakeholder groups helped identify where pockets of data were located, and a presentation was made to our management committee highlighting the benefits a taxonomy could bring. The result was the appointment of Alice in October 2022 to take the lead on this.

After initial inductions we spent the first few weeks meeting with key stakeholders in the firm, including Business Development, HR, Risk, IT and Accounts, all of whom have ownership over key systems and data which will be involved. After these meetings we identified three main strands in the firm where adding a taxonomy would add value to the data:

- Knowledge – precedents, knowhow and internal credentials (who does what)
- Business Development – pitches, sectors, marketing, legal directories, intranet/internet
- Business operations – pricing, people development and resourcing, strategic focus

At this current time Business Development were deemed to have the greatest need for assistance, so this is where we have started.

Howard Kennedy's Business Development (BD) team need to leverage the power of a taxonomy in their work capturing examples of assignments the firm has undertaken for use in promoting expertise to others. The team has developed a consistent way of capturing the data for these assignments across the firm's departments – the use case will be in pitches and credential documents, on the website and for others to understand who have relevant experience. The Business Development team asked us, the Knowledge team, for help in determining the values that should be used to describe the assignments to ensure consistency of description and to improve search and retrieval. The fields needing to be populated were: Client type; Jurisdiction; Country; Sector / Sub-sector; and Work type / Sub-Work type.

The Knowledge team immediately submitted the data from the noslegal taxonomy places facet for use as values in the Jurisdiction and Country fields (more on noslegal later). This was a well-received quick win. The data is extensive, externally reviewed and updated so can be used immediately.

We put together a list of client types, with description and examples. These are currently under review. They may be difficult for users to manually assign.

After initial discussions about sectors with the BD team, our general strategy was to follow noslegal and choose a single existing up-to-date, regularly maintained sectoral taxonomy as a starting point. We chose Statistical



Thomas Jefferson
3rd U.S. President

Thomas Jefferson was an American Founding Father, the principal author of the Declaration of Independence, and the third President of the United States. [Wikipedia](#)

Born: April 13, 1743, Shadwell, VA
Died: July 4, 1826, Charlottesville, VA
Presidential term: March 4, 1801 – March 4, 1809
Spouse: [Martha Jefferson](#) (m. 1772–1782)
Party: [Democratic-Republican Party](#)
Awards: AIA Gold Medal

Get updates about Thomas Jefferson

People also search for [View 15+ more](#)

 John Adams	 George Washington	 Benjamin Franklin	 James Madison	 Alexander Hamilton
---	--	--	--	---

[Feedback](#)

Google, CC BY-SA 3.0 <<https://creativecommons.org/licenses/by-sa/3.0/>>, via Wikimedia Commons

Wikipedia uses taxonomies to suggest further reading⁴

Classification of Economic Activities in the European Community (commonly referred to as NACE), as the closest fit to our requirements. As we worked with BD on the labels for sector terms, we discovered that, though the basic structure of NACE worked well, the labels for terms often needed to be altered to reflect terminology used in legal services. After further discussions, writing definitions and adding examples, the sector and sub-sector lists are now ready for test implementation.

We are currently creating the values for work types. This is the most challenging and interesting set of values to put together as the view of fee-earners is critical in drilling down to all the firm's work types. The professional support lawyers are invaluable in drafting lists of terms and liaising with fee earners.

In due course all the values should be able to be extended beyond the current manual credentials project, to CRM (client relationship management) and then into our other firmwide systems, most notably our matter inception and finance systems.

INTRODUCING NOSLEGAL

Wanting to create a new taxonomy that meets the needs of your firm doesn't mean you have to start from scratch. There are many existing standards and taxonomies that you can build on while customising in places where terms clearly diverge, a few of which we've already touched upon.

noslegal is the brainchild of Graeme Johnston, the founder of Jualio and a former partner at Herbert Smith

Freehills. It is an opensource taxonomy project and is a voluntary, not-for-profit organisation. However, some major law firms and technology companies are supporting their work by providing financial sponsorship or/and actively participating in the working groups. Howard Kennedy became involved in the project in 2022 and it has been heartening to see the collaboration that can exist amongst like-minded individuals across, what are essentially, rival firms to improve the legal process for all.

The initial vision back in 2020 was to release opensource taxonomy materials based around three general topics: *legal places*, *legal work*, *legal contexts*. Several work streams were put together to formulate initial materials with a view to make an initial release in February 2022. As time progressed it was decided to focus specifically on legal places and work as that seemed to be where most benefit could be derived. The aim was to provide a high-level framework based on simplicity and that wasn't tied to a particular legal system. Many projects fall at the first hurdle when they get bogged down in the detail.

The first release of noslegal was in March 2022 on Github with four facets: *places*, *work*, *subjects*, and *perspectives*. Gradually, as word spread, more firms and organisations joined the initiative and a second major release of the project took place in May 2023. As well as updates to the existing facets there was the addition of five new ones: *needs*, *sectors*, *laws*, *information assets* and *combinations*. This time there was an additional slide deck from a workstream led by Kara Redmond of Shearman and

Sterling (which we were pleased to assist with) regarding implementation of taxonomy. Even when the building blocks are provided, we found the next big challenge is applying it to your own internal needs! For anyone interested in getting started with their own firm taxonomy we highly recommend looking at this excellent resource along with the release notes and implementation deck. Details on how to view on Github can be found on the project website: <https://www.noslegal.org/>

Of course, there are many other existing legal taxonomies out there including SALI (Standards Advancement for the Legal Industry)⁹ and Thomson Reuters' legal taxonomy. noslegal doesn't seek to replace these, and all can be options when looking at expanding your taxonomy into more specialist areas.

CONCLUSION

Taxonomies are often viewed as something old fashioned in the brave new world of AI. The reality is that they will still be the foundation that underpins much of the new technology. Taxonomies can link information in all sorts of ways across different communities by getting people to speak the same language.

We will leave you with another old phrase: "sometimes the old ideas are the best". Most new innovations come out of old ideas and the success of this new technology is going to depend on good, quality, structured data. If there's one thing we can be certain of: taxonomies aren't dead yet.

Footnotes

¹ <[https://en.wikipedia.org/wiki/Death_and_taxes_\(idiom\)](https://en.wikipedia.org/wiki/Death_and_taxes_(idiom))>

² Kathryn Armstrong, 'ChatGPT: US lawyer admits using AI for case research' (BBC News, 27 May 2023) <www.bbc.co.uk/news/world-us-canada-65735769>

³ <www.england.nhs.uk/digitaltechnology/digital-primary-care/snomed-ct/>

⁴ This file is licensed under the [Creative Commons Attribution-Share Alike 4.0 International](https://creativecommons.org/licenses/by-sa/4.0/) license.

⁵ Alan Turing Institute

⁶ <<https://medium.com/openlink-software-blog/what-is-dbpedia-and-why-is-it-important-d306b5324f90>>

⁷ <<https://enterprise-knowledge.com/why-a-knowledge-graph-is-the-best-way-to-upgrade-your-taxonomy/>>

⁸ <<https://ironcarrot.com/>>

⁹ <www.sali.org/>

Biographies

Alice Laird is Information and Metadata Manager at Howard Kennedy. She is involved with enquiries and research and works on a range of taxonomy projects to improve data consistency and information management. She previously worked at ICAEW, government departmental and Lambeth Palace libraries in a range of metadata, systems and customer service roles.

Katy Snell is Knowledge Systems and Information Manager at Howard Kennedy. She manages the online research subscriptions for the firm as well as carrying out enquiry work and assisting in knowledge projects including document automation. She previously worked at ICAEW and Ministry of Justice libraries in systems and customer service roles. She is a member of BIALL Council.