SWIFT INSTITUTE

SWIFT INSTITUTE WORKING PAPER NO. 2012-005

The Prospects for Common Language in Wholesale Financial Services

MALCOLM CHISHOLM

ALISTAIR MILNE

PUBLICATION DATE: 9 SEPTEMBER 2013

The Prospects for Common Language in Wholesale Financial Services[1]

Malcolm Chisholm[2] and Alistair Milne[3]

September, 2013

Abstract

This paper examines the idea of a common financial language (CFL) from both conceptual and practical perspectives. There can be no single CFL because there is no single underlying and unchanging financial reality. Potential efficiency and risk-management benefits from developing greater CFL for data management are large; but in practice CFL must be confined to a limited number of shared concepts, with fuller agreement on definitions and relationships for use by particular 'communities of interest' within firms and in specific operational processes. CFL will be a process of gradual evolution and adoption, not a once and for all linguistic revolution.

Keywords: Computer standards, Database management, Enterprise data, Financial Definitions, Financial Ontology, Financial Risk Management, Operational efficiency in financial services, Regulatory data

JEL numbers: G20, G28, M15

1. Introduction

The call for a 'common financial language' (CFL) is now a topic of much discussion in regulatory and industry circles. From the regulatory side, a speech by Andrew Haldane delivered to SIFMA (Securities Industry and Financial Markets Association) in 2012 has had a powerful impact ((Ali, Haldane, & Nahai-Williamson, 2012)). This speech 'Towards a common financial language' is perhaps the best-known example of a description of what a CFL is and the benefits it would bring. Specifically Haldane refers to current initiatives for developing global legal entity identifiers (LEI) for wholesale market participants and suggests that these LEIs could be the 'nouns' of a common financial language with an accompanying system of product identifier's (PIs) the corresponding 'verbs'.

[1] This research has been supported by a grant from the Swift Institute. We thank Stijn Christiaens, Stephen Lindsay, and members of the Swift Institute's Advisory Council for comments.

[2] AskGet consulting: www.askget.com; masterdataconsulting@googlemail.com

[3] Corresponding author: Loughborough University, School of Business and Economics, a.k.l.milne@lboro.ac.uk

On the industry side, there have been several efforts to develop more common language. One of the best known, the most ambitious and the closest to the issues which we examine in this paper is the Financial Industry Business Ontology (FIBO).[4] This industry initiative, begun in 2008 and stewarded by industry group the Enterprise Data Management Council (EDMC), aims to develop a 'data ontologies' for financial services i.e. a conceptual mappings of instruments, entities and relationships together with tools and techniques to make these concepts operational in computerised processing. In the FIBO project EDMC is working with the standards consortium, the Object Management Group, to create a range of such ontologies.[5] This effort has begun with some specific financial instruments (the initial 'proof of concept' has been on interest rate swaps) and it is hoped that FIBO will ultimately to cover many other if not most aspects of financial services.

This paper examines the prospects for development of such common language. This is a goal well worth pursuing, but our examination suggests that establishing a single common financial language for all wholesale market activities may be an unrealistic goal. We find rather that establishing common financial language will be an evolutionary process towards more shared definitions and understanding, not a task that can ever be entirely completed.

We develop this argument as follows. The first part of the paper (Section 2) is conceptual, considering what is or can be meant by a common financial language. It interprets this challenge as being, at least primarily, about establishing common financial language for the exchange and summary of information held in computer data systems. It then considers the wide range of different views about the relationship between language and underlying reality expressed over the centuries in the philosophy of language, abstract issues but of considerable relevance to computer system design. Finally it distinguishes different aspects of common language for communication between business computer systems: (i) financial glossaries and dictionaries; (ii) communication 'standards'; (iii) identifiers and (iv) 'ontologies' (such as those being developed by the EDMC).

---

[4] The progress on FIBO can be followed via http://www.edmcouncil.org/financialbusiness . As described on the EDM council website "FIBO is a collaborative effort among industry practitioners, semantic technology experts and information scientists to standardize the language used to precisely define the terms, conditions, and characteristics of financial instruments; the legal and relationship structure of business entities; the content and time dimensions of market data; and the legal obligations and process aspects of corporate actions."

[5] The Object Management Group (OMG see www.omg.org) is an international computer industry standards consortium that supports open standards to facilitate 'enterprise integration'

The following two sections focus on some practical aspects of common financial language, considering how common financial language might work in the context of automated computer processing and database management (Section 3) and examining the business or economic benefits it has offered in the specific contexts of communication between financial market participants and in entity identification (Section 4).

Section 5 concludes, arguing that such common language cannot be entirely context free – there can be no single vocabulary that meets all needs of communication and record keeping. The prospect is therefore for evolution towards greater commonality of language rather than the creation of a single common financial language.

The Appendix, supporting the discussion of the concept of a common financial language in Section 2, provides an examination of the functionality of a widely used finance dictionary, the Campbell R. Harvey Hypertextual Glossary.

2. What can be meant by 'common language' in financial markets?

This section argues that establishing such common language is primarily about improving communication between computer systems. But the notion that there can be a single universal and unambiguous financial language, for exchange of information between computer systems, adequate for capturing all aspects of financial transactions and financial services is misguided – the underlying reality is too complex and too fluid for such a construct.

Different types of common financial language

There already is a great deal of shared language used by human participants in wholesale financial markets. The existence of this common language does not mean that there are no problems of understanding, but humans are used to coping with ambiguity and typically misunderstandings are recognized and corrected relatively quickly. For day to day communication between financial market professionals there does not seem to be any serious problem of lack of common financial language.

This is not to say that financial language used by human beings is without shortcomings. There can be conceptual differences that undermine common understanding and can create financial fragility. To take one prominent example from the financial crisis, the concept of ‘credit rating’ was fatally weakened as the ratings were applied to complex structured products as well as to corporate and sovereign bond issuers. In the case of corporate and sovereign issuers there was a clear underlying judgement based system for assessing the borrower strength. Individual ratings could be challenged; but, to use the statistical term, they were unbiased; i.e. ratings mistakes whether favourable or unfavourable averaged each other out; and they were continually subject to challenge from market professionals

In the case of structured products, investors mistakenly assumed that A or BBB meant the same for the tranche of a structured security as for a corporate bond issue, even though the mechanism for assessing ratings was entirely different. These structured credit ratings meant something quite different than the corresponding corporate or sovereign ratings, for at least three reasons: (i) the reliance on statistical rather than judgement based methodology meant that these ratings could not be easily challenged by investment professionals; (ii) the undermining of quality control resulting from the practice of ‘issuer pays’ combined with the fact that so many structured securities were issued by the same few banks who could threaten to move their lucrative business elsewhere if they did not like the ratings; and (iii) as we now know at least in the case of sub-prime mortgage backed securities and ABS-CDOs a further ratings bias introduced by gross underestimation of the degree of default correlation in US housing markets.

This example indicates that in one important respect a misunderstanding of language was an important contributor to the global financial crisis. But the problem was not something that could have been immediately fixed by a change of vocabulary; rather the problem was the flawed conceptual framework underlying the language of ratings and its application to the tranches of structured credit problems, together with the incentive problems that encouraged issuers and investors alike to ignore these fundamental difficulties.

Another area where the shortcomings of financial language are exposed is in law and regulation. A term such as interest rate swap, commodity future or covered bond may be well understood by market participants, but when it comes to enforcement of contractual obligations a much more precise language that will survive scrutiny in court of law is required. Thus all financial contracts and instruments are accompanied by extensive supporting legal detail. Particular challenges arise from differences in legal language and concepts in different jurisdictions, something highlighted by a recent European Commission examination of financial terminology that examined more than one thousand terms used in EU financial regulation.[6]

Reviewing more carefully both the Haldane speech and the goals of the FIBO project, it seems clear that what they have in mind is not a common language for communication between human beings, but rather a common and standardised language for automated processing of data. In Haldane's case he is reflecting the desire of regulators that individual firms should be in a position to effectively aggregate their exposure data, and analyse this data across a range of different dimensions; and that regulators should be able themselves to collect and aggregate exposures across firms in order to monitor systemic financial risk. Standardised identification of both counterparties and of instruments would seem to be a precondition, but it is not entirely accurate to describe these as the 'nouns' and 'verbs' of a common financial language. They are rather classification keys that can be used to understand and manipulate granular data.

FIBO has a similar goal, to provide a detailed framework for recording data that can support automated processing and meaningful manipulation and aggregation. It though goes further than Haldane by seeking to have standardised parsing of the contractual components of different financial instruments and of the relationships between different legal entities. To the extent that this can be done in a way that can be readily utilised by financial markets, then it will indeed provide a common and standardised financial language that will support both automated and low cost business processing, and convenient aggregation of risk exposures from granular data, at both firm and system wide level.

In summary: the proposal for common language put forward by Haldane was not aimed at changing the concepts used in risk or other reporting, at resolving legal disputes, or at reducing the many potential misunderstandings in the constant exchange of telephone and email messages between market participants His illustrations of the benefits of common business language (the barcode used in retail, supply chain and healthcare; the world wide

[6] (European Commission, 2013). A well known retail financial example of such inconsistency is the direct debit, the legal and regulatory arrangements for which different quite dramatically from one country to another (for example the process for establishing or cancelling the arrangement and the rights of the payor to get their money back if the debit has not been properly authorized). These inconsistencies have blocked efforts to establish an EU wide direct debit arrangement.

web) are all about the more effective summary and exchange of information held in computer systems and databases. In this paper, therefore, we focus on the prospects for a common financial language that can be used for communicating between computer systems.

Lessons from the philosophy of language

As a precursor to this task we first take a step backwards, and ask the more general question, what exactly can we mean by ‘common language’, whether in communication between computer systems or in any other context? Discussions of this question have a lengthy intellectual pedigree, with roots that can be traced back at least as far as the classical Greek philosophers.

Two of the most influential perspectives on the way language works are those of Plato and of Aristotle. Plato’s most prominent view of the philosophy of language is his theory of ‘forms’, discussed at some length in three dialogues (the *Phaedo*, the *Republic* and the *Meno*) and more briefly in other dialogues.[7] Most famously in the *Republic* Plato discusses language and education, drawing an analogy with prisoners chained in a cave who observe, and talk about, shadows of objects which they see thrown onto a wall in front of them by the light of a fire. One prisoner is freed and is able to see the real objects in sunlight and then returns to the cave. But he is unable to persuade the remaining prisoners that the shadows are not real. According to Plato behind our perceptions (the shadows perceived by the prisoners) and the words that we use to describe them there is an underlying reality (the ‘forms’) which philosophy and education can help us understand (just as sunlight helped the escaped prisoner sees the true objects that cast the shadows).

Aristotle’s major contributions are from his logic in which he considered the way in which arguments are made.[8] The core of Aristotle’s logic is deduction: understanding how some assertions can be derived from others. According to Aristotle propositions, statements which can be true or false, always consist of a subject (about which the proposition is made) and a predicate (an assertion about the subject). He also argues that definitions, which are a form of proposition, can be understood in terms of an ascending hierarchy: of individual, species and genus. An individual, whether an object or a living thing, has characteristics (or differentia) associated with its species. A popular way of presenting these hierarchical

[7] See (Crivelli, 2008) for more detailed discussion.
[8] See for example (Smith, 2000)

relationships especially amongst medieval logicians was as a so called Tree of Porphyry, a visual representation of these relationships in which the ladder from the highest genus the 'summa genus' down to the lowest species represent the main trunk, and various branches indicate the 'differentia' that distinguish each level.[9] A definition of the species can be asserted by stating the genus to which the species belongs and its distinguishing characteristics (a well known example although not actually in Aristotle's writings being that the species human is the member of the animal genus with the power of reasoning).

A great deal has been written and said about the philosophy of both argument and of language since the time of the ancient Greeks. Formal logic has developed much more precise analyses of how arguments can be made, applicable in a range of contexts, for example what is a correct mathematical proof (although much of Aristotle's less formal discussion of logical argument is still valid). The modern philosophy of language has explored the idea that language can be broken down into constituent parts and in order to assess its validity or truth; that statements can be divided into those that are valid by virtue of logic and those that can be empirically distinguished as true or false; and hence that language can be fully understood from a knowledge of its constituent parts and the truth or falsehood of individual statements.[10]

During the twentieth century this view – associated with Frege and Russell and the school of logical positivism – came to be regarded as oversimplified. No single view of the philosophy of language has replaced it; instead there are a variety of different perspectives, including viewing language as a social act (i.e. it is not possible to separate language from its social context and meaning); that language and expression are closely tied up with the issue of verification (i.e. what are the processes for deciding if statements are true or false); and scepticism on whether there really is an entirely clear distinction between logically true and empirical statements.

The most influential but also enigmatic contributor to these debates was Ludwig Wittgenstein. In his early work the Tractatus he engages, through a lengthy series of propositions and propositions on propositions, in an attempt to establish a complete structured analysis of language and knowledge in the tradition of Aristotle, Frege and

[9] http://en.wikipedia.org/wiki/Porphyrian_tree
[10] (Wolf, 2009) provides an overview.

Russell.[11] Later, after a lengthy break from the academic work and then a return to the University of Cambridge, he spent several years working on a radically new perspective, eventually published posthumously as the Philosophical Investigations and in various transcribed lecture notes and preliminary studies.[12] Here he turned away from the idea of an underlying logical structure of language emphasizing instead the notion of 'language games', a diverse set of rules and arrangements that depended on context and application.

This brief – and extremely shallow – discussion of logic and the philosophy of language may seem like an excursion away from the subject at hand: 'common language in wholesale financial services'. But these on-going philosophical debates offer important practical lessons and as a result are part of the common discourse of both researchers and practitioners in information and data management.[13]

They offer a number of lessons for the 'common language' debate. One is that Haldane, in referring to the 'nouns' and 'verbs' of a common financial language, adopts a position much like that of Aristotle, Frege or Russell implying that if we only get the elements of language pinned down, then the central problems of understanding of and communicating financial exposures can be resolved. This may be true in some situations, but we need to be aware of the great variety of ways in which language, including financial language, is used and hence that agreed definitions and accuracy of terms may not always be achievable, without substantial changes in business practice; or if it is achieved may not resolve all problems of communication.

Another is that, financial language is inherently complex because financial terms refer to an underlying social, economic and legal context; the future returns to a loan exposure for example will depend upon the characteristics of the borrower, the power to legally enforce contractual obligations, and also on future states of the economy and outcomes for market prices. The meaning and interpretation of financial terminology depends on the understanding of other financial terms in a lengthy and possibly even circular series of

---

[11] (Wittgenstein & Russell, n.d.), a translation of the German edition originally published in 1921.

[12] (Wittgenstein, 2009) is the latest edition of the original 1953 publication.

[13] There is an extensive literature applying a wide range of concepts from western philosophy to practical problems in data and information management. One influential writer is John Sowa (see (Sowa, 2000)) whose graphical representations of logical relationships, drawing on the work of the 19th American pragmatist philosopher Charles S. Peirce, have been taken up by many analysts. We prefer not to go into these further complications.

iterations: a credit rating for example depends on a quite complicated assessment of various other aspects of financial reality. Moreover, as we discuss further below, all these underlying institutional and business arrangements can change substantially over time and from one jurisdiction to another.

A third lesson, harking back to the insights of Aristotle, is that we must take account of the relationships between different concepts and levels of reality, for example Aristotle highlights the relationships between the general (genus), the more specific (species) and instances (individuals); but these are far from being the only possible such 'semantic' relationships.[14] A fourth is the need to be aware that, even where language is shared, establishing the accuracy of statements in any language can be challenging and subject to varying interpretation. What is acceptable as verification may depend on the context in which the statements are used.

This does not mean that common financial language is impossible, but that it will always be a work in progress, with room for further improvement and clarification.

An examination of the notion of common language in financial services

With this philosophical background in mind, we now look critically at some of the presuppositions that support the arguments of both the Haldane speech and the FIBO ontology project.

In the recent proposals for common financial language it has not been made clear if it is supposed to describe the reality of the entire financial sector, or if it should be restricted to aspects that are of interest only to certain groups – for example regulators, back office systems, portfolio managers, auditors, or risk managers.

Another possibility is that this language should refer to aspects of finance that are common across a global financial community. This might include concepts that are foundational to finance, and/or concepts that are generalizations from all financial activities.

In the Haldane speech there is no notion that 'finance' may mean different things to different communities. The same is true of FIBO, which has a stated overall goal to:

[14] Pierce's pragmatist logic for example introduces a further aspect, that of interpretation.

> *"standardize the terms and definitions of all reference data stored in master files of financial institutions (unambiguous shared meaning)"* [15]

It seems that there is a presupposition by both Haldane and the FIBO project that a CFL will reflect the existence of a single underlying reality, which is termed 'finance'. The idea that there is a single underlying reality which is to be understood in any domain of experience is a common theme in the natural sciences. It therefore seems plausible that it should apply to finance.

However there are objections. First, finance changes over time. For instance, the advent of asset securitization is relatively recent, e.g. the earliest known securitization of auto loans in the US was in 1985.[16] Prior to this time, securitization of auto loan assets did not occur, but after this time it did. Furthermore, the practice of securitization of auto assets in the US has evolved since 1985. One of the authors of this paper (Chisholm) has personally developed securitization systems for major US firms, and has experienced how features such as credit enhancements and liquidity facilities were added later in time to these systems. This is indicative of the development of new concepts that did not exist before, and which are more akin to inventions rather than discoveries of facts.

Still, even if it is accepted that finance changes, it can still be argued that at any point in time there is still a single underlying reality. It is merely a matter of ensuring that the common financial language is general enough to covering the new aspects of the underlying reality as these are developed. But this runs into a deeper objection that different interpretations of reality are possible at a single point in time. The experience of Chisholm with asset securitization furnishes an example. He participated in the development of systems to manage securitization that were used by every major US credit card issuer. Each such issuer has created various deals over time, resulting in the issuance of bonds to securitize the revolving debt in the pool of securitized credit card receivables.

In one instance, Chisholm was working on a deal of a very large US credit card issuer. *Finance Charges* were calculated based on monthly interest received and a few other factors.

---

[15] ("Understanding the Legal Foundations of Business Entity Types, Ownership and Control," 2012)

[16] Hearing before the U.S. House subcommittee on Policy Research and Insurance in "Asset Securitization and Secondary Markets" (July 31, 1991), page 13, quoted in Wikipedia article on securitization, http://en.wikipedia.org/wiki/Securitization, accessed 2013-02-10

He was then instructed to include *Monthly Overlimit Charges* into this calculation. This was done, and the value of *Finance Charges* increased by several hundred thousand US dollars per month. Thus, in this deal, *Finance Charges* meant one thing at one point in time, and a different thing at another point in time. Also a potentially important risk indicator was being overlooked, because monthly over limit charges that might have been a useful indicator of future credit concerns were conflated with other financial charges.

This example suggests two possible points of view on common financial language. The first is that there were two (or more) interpretations of underlying reality at a single point of time, and the language used to describe the exposure switched from one interpretation to another. Under this interpretation financial language can be relatively straightforward and fit well with the immediate needs of the user, but rarely will it be common to all business situations. The second is that there was really only ever one accurate interpretation, but the language used to describe the exposure was insufficiently granular. This though indicates that a common financial language that captures all realities, even at a single point of time, may have to be extraordinarily complicated.

Another related objection is that when a significant business or regulatory change occurs in one jurisdiction, it does not mean that it occurs in all jurisdictions simultaneously. For instance a Deloitte study in 2007 of the securitization market in South Africa discusses the machinery that enabled sub-prime mortgages to be securitized in the US and Europe, but notes that:

> *"we do not expect to see any contagion effect on the South African market because there are no deals in the South African market that fall into the sub-prime mortgage category."*[17]

From this it can be clearly seen that sub-prime mortgages were not an element of finance in South Africa in 2007. Therefore, the parts of a CFL describing sub-prime mortgages would not correspond to the reality of finance in South Africa in 2007, but would correspond to the reality in other jurisdictions in 2007, such as the United States.

[17] (Deloitte, 2008)

From this we can see that a major difficulty with the notion of a common language describing a single underlying reality for finance is the existence of differences between contexts in which finance is carried out. A single reality implies something that can be treated as a whole, and which consists of parts that themselves can be dealt with individually. An alternative – a widely used perspective in the social but not the natural sciences – is that the reality of finance is a system formed by the interaction of many independent entities in an ever changing system of relationships. F. A. Hayek, a proponent of this alternative view describes the subjects of all social sciences as "an organization rather than an organism".[18]

There are several different contexts that need to be considered. One is market practice. For instance, AmBank Group undertakes the practice of Islamic Banking in Malaysia and has posted a number of Islamic Banking concepts with definitions on its website.[19] One example of such a concept is as follows:

> ***Qardhul Hasan (benevolent loan)***
>
> *Refers to an interest-free loan. The borrower is only required to repay the principal amount borrowed, but he may pay an additional amount at his absolute discretion, as a token of appreciation.*

This term could not be found in the FIBO semantics repository.[20] Neither could a definition that approximates to the one on the AmBank site.

Another is differences in law and regulation between jurisdictions. For instance, Solvency II is a pan-European insurance regulation initiative which has defined new concepts. One such new concept is data credibility. This is described as follows:[21]

> *"the undertaking must be able to demonstrate that it recognises the data set as credible by using it throughout the undertakings operations and decision-making processes."*

What this means in practice is that insurance companies must use data that is present in their operational systems as the basis for their reporting to regulators. There must be a

[18] (Hayek, 1980)
[19] http://www.ambankgroup.com/en/Islamic/AboutUs/Pages/BankingConcepts.aspx last accessed 2013-02-10
[20] http://www.edmcouncil.org/semanticsrepository/index.html accessed 2013-02-10
[21] (CEIOPS, 2009) *Paragraph 3.66*

direct lineage of data from the operational systems to what is presented to the regulators. This concept does not appear in US regulation, such as the Dodd-Frank Wall Street Reform and Consumer Protect Act of 2010. Nor does data credibility appear in the EDM Council list of data quality dimensions.[22]

The great variety of different activities at a detailed level within financial firms present another set of contexts, one that is possibly very common. Chisholm's experience with asset securitization (described above) furnishes an example of this. In one instance, let us call it Deal A, *Finance Charges* were defined as *Monthly Interest Received*. The term *Finance Charges* was defined as such in the legal documents underlying the deal. In a separate deal, Deal B, *Finance Charges* were defined as *Monthly Interest Received* plus *Annual Membership Fees* - again in the legal documents underlying the deal. The bonds of both these deals coexisted in the market at the same time. *Finance Charges* were reported as numerical quantities to the investors for both deals, and also to the SEC's Electronic Data Gathering and Reporting (EDGAR) system. However, the underlying definitions were different in each case.

Many other examples of financial terms differing according to context can be produced, including variations amongst:

- functions within firms. Thus in the context of the broker-dealer activities of investment banks, there may be terms and definitions (language) that are specific to customer *execution* and different terms used in the context of proprietary *trading*. Other for *post-trade processing*, another for *audit* (mark to market), yet more for different aspects of risk-management including *compliance*, *market risk* reporting and management, *counterparty risk* management, *operational risk* assessment, *treasury* and balance sheet funding. A number of functions e.g. collateral management, data and information systems, legal advice on contract definition and contractual disputes, can bridge more than one aspect of the firm. Every one of the functions can develop its own separate language.

[22] http://www.edmcouncil.org/content/view.cfm/downloads/20120712_Data_Quality_Dimensions.pptx.pdf , last accessed 2013-02-10

- industry segments, thus as well as broker dealer activities, we might distinguish *corporate advisory* work, which could include M&A and primary market deals*, corporate banking,* household *retail* banking and also small business retail banking, *private banking*, include wealth management and family banking, *private equity* including venture capital, *asset management* for retail clients, for institutional investors and for hedge funds, *casualty* insurance, *general* insurance and pension funds. A number of specific vocabularies have developed within these segments.
- sub-markets within industry segments, so for broker dealers it is usual to distinguish *fixed income* and fixed income derivatives, *equities* and equity derivatives, *forex* and forex derivatives, *money markets* and *commodities.* This list is not complete, for example this list does not include exotic derivatives and does not distinguish product innovation. Further breakdowns also arise for other industry segments.
- from one firm to another. The same terms may have different meanings or different terms have the same meaning when comparing one firm with another. This applies especially to internal functions such as audit, risk management, data and information systems generating extensive challenges when firms co-operate on joint activities or for the process of post-merger integration. It can also apply in the context of trading between firms, where for example mark to market valuation of the same position is not always carried out on a consistent basis between firms.
- from one jurisdiction to another, so the same term (or dictionary translation of a term) may have a different meaning; or different terms may have the same or a similar meaning, as a consequence of legal, cultural, regulatory and historical differences.

Even if is accepted that variations exist, perhaps these can be accommodated within an overall common language that sees them as parts of a whole. But these different contexts that have been described are not parts of finance in the sense of being discrete parts of a whole (in the sense that for example England plus Wales plus Scotland constitute the island of Great Britain). If that were the case, then they could each be described and the sum of them would be a single reality of finance.

Rather, these contexts:

(a) Deal with general concepts in a different, specific way compared to other contexts.

(b) Consist of different systems of relations.

(c) Have unique concepts and relationship systems.

The specificity in (a) and (b) is not truly unique. It may be shared with some other contexts. For instance in the example of *Finance Charges* including *Annual Membership Fees* stated above, a number of deals included *Annual Membership Fees* in *Finance Charges*, while others did not. What is different about each context cannot be understood from a general description or definition of the context. Rather, it must be determined by careful investigation of the context itself.

This supports the view that there is no single underlying reality of finance, but rather a set of different realities in different contexts.

Different aspects of common business language

To complete this section on the concept of a common financial language, we briefly discuss some different aspects of common languages used in business, especially for communication between computer systems. These are: (i) terminological dictionaries; (iii) communication standards; (iii) reference identifiers; and (iv) data and information ontologies.

One interpretation of the idea of a common financial language is that it is a single set of definitions – a kind of über-dictionary of financial terminology. Neither Haldane nor FIBO go so far as to endorse such an interpretation. Haldane is more concerned with the problems of ambiguity of identification, both of legal entities and of financial instruments. FIBO are careful to draw a distinction between a conceptual ontology i.e. a hierarchical categorization of financial concepts – and the implementation of this ontology in automated systems.

Still it is useful to consider how far such a standardized dictionary might take us towards a common financial language. The Appendix to this paper reports the outcome of an examination of the functionality of the most widely used financial dictionary; the Campbell

R. Harvey Hypertextual Financial Glossary. This finds that the functionality of this dictionary, is rather limited, not doing a much deal more than satisfying the need for an English definition of a term that is already known. It does not provide the kind of detailed analysis of the relationship between terms that is sought for example in the FIBO ontologies.

Dictionaries such as Campbell R Harvey's are useful in the context of human communication, for example as an 'aide memoire'. Human beings manage quite well with ambiguity, they can guess successfully at meanings from context, they may not recall information as accurately as machines, but they can use machines to support their memories and they can recognise relationships amongst different categories without this needing to be fully spelt out. In the context of human interaction or human-computer interactions, it can certainly be useful to have available a dictionary of financial terms, for example via a terminal such as Bloomberg where the Campbell R Harvey dictionary is provided, that can be consulted before making a verbal communication or a manual entry into a computer or trading system. Also, as we will discuss also below, lists of definitions play an important role in the implementation of both communication standards and in data and information ontologies.

Standards – agreements on ways of doing or making things – play a major and often underappreciated role in business activities.[23] To take one example technical engineering standards (of which there are very many) are needed in order for computers or other information technology hardware to communicate data to each other. Standards are also critical to ensuring that different products work together (a basic example is the standardization of screw threads, but there are many, many others). Increasingly standards have also been developed for business processes, such as human resource management or quality control, as well as for individual products and services.

Over time a variety sophisticated institutional arrangements have grown up for developing, formalizing and promoting business standards. Most countries have their own national standards bodies. These co-operate through the International Standards Organisation (ISO) and in addition there are a number of national and global standards organizations operating within individual industries.[24]

[23] For reviews of the economic role of standards see (Blind, 2004; Swann, 2000, 2010)
[24] A useful review of the work of ISO is provided by (ISO, 2010)

How do business standards relate to common financial language? In order for data to be interpreted it must not only be transmitted, it must also be organized in agreed ways i.e. there must be communication standards. These communication standards (Section 4 below lists several of these) are in effect common languages already in widespread use in financial markets. Still such communication standards, while essential to operation of many financial markets, are not universal and they do not deliver all the benefits of common financial language as conceived by Haldane. For example, even with further development, improved communication standards between firms will not help regulators understand better risk exposures or threats to the financial system as a whole.

More relevant to the better management and monitoring of risk is reference identification, both for entities and for financial products. The financial services industry has developed a number of such identifiers, for example ISIN (international security identification numbers, which incorporate a number of different identifiers such as US CUSIP), BIC (Bank identifier codes) and Dunn and Bradstreet numbers. In addition many firms maintain their own internal identification systems for identifying counterparties or financial contracts. These and other identifiers play an important role, in the execution of financial transactions, in regulatory compliance and for aggregation of information. They remain however fragmented and inconsistent, so while they are part of the financial language used within and between firms, they fall short of being an entirely common financial language. Haldane's speech (Ali et al., 2012) rightly stresses the importance, for better risk monitoring and management, of improving on this patchwork of different identifiers by establishing unique and universal referencing that can be used across the financial system. The global legal entity identifier system (discussed further in Section 4 below) is a first effort of this kind.

Yet another aspect of common business language are 'ontologies': agreement on the terms used by computer systems so that they can better work together. The FIBO ontology being developed by the EDMC is one example, but it is far from being the first or only business ontology. There has been an explosion of work on the application of ontologies in practical problems of computing and information management; on general methods for the construction of ontologies; and on ontologies as a sub-discipline of business, computer science and information management research.[25] Prominent amongst these has been the promotion by the World Wide Web consortium – the standards body for the internet -- of ontologies as part of its effort to create the 'semantic web' i.e. to develop internet to allow data to be 'shared and reused across application, enterprise, and community boundaries' and the development of a language (OWL) for the expression of both ontologies and underlying knowledge.

As described by (Staab & Studer, 2009) ontologies are 'a formal description of concepts and relationships that can exist for a community of human and/or machine agents.' As such, ontologies are a form of common language for exchange of information between computer systems. An ontology can be successful in facilitating such exchange by providing a categorization and structure of concepts which bridges the different contexts in which individual systems store and organize information.

It is important to understand the limitations of ontologies. (Hepp, 2008) provides an insightful and critical review. He notes that the term ontology is itself not a clearly defined and there are often substantive differences in the construction and application of ontologies: for example whether there is a formal logical expression (e.g. as in the tree of poryphry) of the relationships between concepts in the ontology or not; whether the ontology represents the underlying relationship amongst concepts or instead is developed as a specification of these relationships in some form that can be interpreted and used by a computer system ('machine-readable'); and whether the ontology is supposed to represent an underlying true set of conceptual relationships, or whether instead it is outcome of a consensus amongst a 'community' of users about concepts and their relationships.

(Hepp, 2008) also discusses some of the important practical issues that arise in the application of ontologies. Unlike a human language, an ontology requires on-going work to maintain it and ensure it can be used across the different systems it is supposed to bridge. There are trade-offs. For example an ontology that serves very large community must necessarily be 'easy to understand, well documented, and of limited size' ((Hepp, 2008) pg. 8). If this view is correct then a single ontology for all of wholesale financial markets is simply not achievable.

[25] (Staab & Studer, 2009) is representative.

Incorporating a more sophisticated formal structure between concepts may allow the ontology to achieve a greater range of different types of communication between computer systems ((Hepp, 2008) describes this as 'expressiveness'), but also make it more difficult to use and understand. Ontologies of finance, such as FIBO, can certainly help improve communication amongst the many computer systems used in wholesale financial markets, and with the humans that use them, but there are important practical constraints and useful financial ontologies are a prospect for specific 'communities' of users rather than for the entire industry. We will discuss some of these practical constraints further in the following Section.

Summary

This section has analysed the concept of a common financial language and found that such a project is more complicated that might be understood from a casual reading of recent pronouncements by global policy makers or industry leaders. The practical reality of common financial language runs head on into some profound philosophical issues. There is no unchanging financial reality that can be described by such a common financial language. The language itself if it is to be meaningful cannot simply be a list of primitive definitions that can be summarised in some large dictionary; rather like any other human language it must be a flexible creation that is to an important degree dependent on context and usage.

Progress in the development of common financial language for communication between computer systems, would seem to consist of a number of distinct but related practical activities. These include development of reference identifiers to support the aggregation of information from different computer systems; agreement on common standards for the communication and recording of data, so that computerised systems can process and exchange information with each other without manual overrides; and the development of business ontologies, i.e. agreement on the meaning of concepts, the relationships between them and their implementation in business systems, in order to support interaction between firms. Progress on these aspects of common financial language can lower operational costs, reduce both financial and operational risks in individual firms and help regulators better monitor and manage risks to the financial system as a whole.

3. Common language and communities of interest

The preceding section has considered the concept of common financial language and concluded that to be of practical value, then this must be about developing common approaches (including messaging standards, data referencing and ontologies) to the communication and organization of data, so that computers and their associated databases can communicate with each other automatically without the need for frequent manual intervention.

This section considers what can be achieved along these lines, from a database management perspective. It argues that there cannot be a single CFL for communication between computer systems and databases; rather the prospect is for development of several interlocking common languages, each tailored to a particular context e.g. on-boarding of new customers, the provision of pre-trade information, trade execution, post-trade processing, risk measurement and management, or accounting and valuation. Specific languages are thus associated with communities of interest – groups of individuals in a firm or in firms that are using the same computer systems for specific business purposes. We find that complete standardization of language across all these communities of interest is either infeasible or very costly (we discuss in particular the costs of adopting CFL due to the reliance on legacy systems). But there can be progress towards greater standardization to facilitate risk management and operational processes such as mergers, divestitures or the administration of corporate failures.

The conflict between database design and common financial language

We can begin by considering the optimization of data store design and its alignment to a common financial language. The most common platforms used to store data today are relational databases. The characteristic of a relational database is that data records are referenced by a *primary key*. For example in a banking context the primary key might be an account number. A relational database can contain many tables of data each with their own primary key. For example an account record might contain a cross reference which is the primary key to a different data table. This cross reference is sometimes called a 'foreign key'. Thus an account record might contain a foreign key that leads to a customer record containing details such as name, birth date and address.

The flexibility of relational databases has led to them becoming the dominant format for storage of data in corporate applications. Examples include Oracle, IBM DB2, MySQL, Teradata, and Microsoft SQL Server.

Given such platforms, the question arises as to whether the way the data is stored in them can be made to correspond to any CFL. This question is not trivial, since, if conceptual progress is made in developing a CFL, but the financial language cannot be practically implemented in commonly used database platforms, then the effort would appear to be in vain. For the purposes of examining this question, we will concentrate on relational database platforms as they are more thoroughly understood than the emerging "Big Data" platforms.

Relational database theory was significantly developed by E.F. Codd, who in 1970 proposed a relational model in which data was organized into tables based on the logic of the interdependence of data elements.[26] This theory introduced the idea of First, Second, and Third Normal Forms, which are rules for the design of relational databases. These ideas were quickly adopted and led to the dominance of relational database platforms which continues to this day.

From this we can already see that irrespective of any financial language, database designers have an internal set of criteria that they use to design databases. This would appear to contradict the notion that a database design can be completely derived from a CFL.

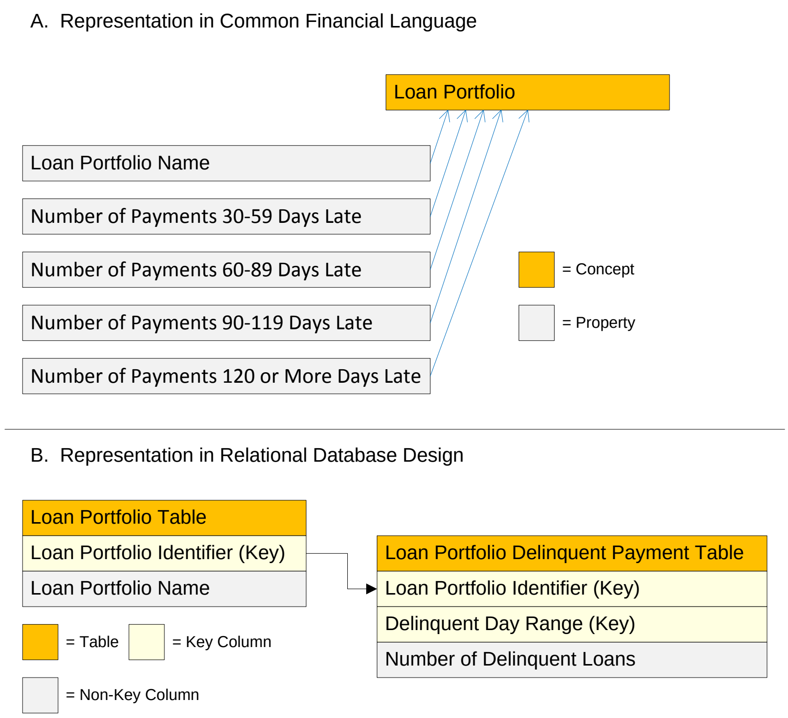
For example, Delinquent Loan Payments in a Loan Portfolio may be classified into:

- Number of Payments 30-59 Days Late
- Number of Payments 60-89 Days Late
- Number of Payments 90-119 Days Late
- Number of Payments 120 or More Days Late.

In a CFL, these would reasonably be expected to be properties of a Loan Portfolio. Loan Portfolio would be a concept, and each of the above data elements would be a property of it. However, in relational database theory, this arrangement would violate First Normal Form which dictates that no relational database table can have "repeating groups". These data elements could not be placed as columns in a Loan Portfolio table. Rather, there would

[26] (Codd, 1970)

have to be an additional table in which the above data elements would be stored as records, rather than as columns. Figure 1 shows the contrast.



C. Example of Data in Loan Portfolio Delinquent Payment Table

Loan Portfolio Identifier	Delinquent Day Range	Number of Delinquent Loans
MBS2008-01	30-59	2307
MBS2008-01	60-89	3112
MBS2008-01	90-119	11234
MBS2008-01	120+	5725

Column Headers

Data Records

Figure 1: *Comparison of CFL and Relational Database Equivalent*

From Figure 1 it can be seen that Loan Portfolio Delinquent Payment Table really has no equivalent concept in the CFL - it is a construct that is required in order to get the database into First Normal Form. Further, we can see that the individual concepts in the CFL for each of the distinct data elements representing delinquent payment "buckets" have been generalized in the relational database design. These data elements do not exist as distinct columns but have to be interpreted from the records stored in the Loan Portfolio Delinquent Payment Table.

The reasons commonly given for creating relational database designs are:

- Reduction in the risk of update anomalies. This is a form of programmer error, whereby the programs contain bugs and do not ensure that related data is updated synchronously.
- Reductions in need to restructure the database as further requirements for it emerge. The database design will only change as the business changes. The task of adding a single column to a production relational database is a complex task due to the risk of causing system problems and is generally actively avoided unless it is really necessary.

As can be easily understood, these reasons are not concerns of any CFL. Again, therefore, we see that the drivers for database design are quite independent of those for a CFL. The technical staff responsible for implementing a database will almost certainly be oriented to the technical drivers of database design, rather than aligned to those for a CFL.

Despite the success of Codd's relational database theory, there is a surprising level of deviation from it in practice, even by professionals. This is because if pure relational design is pursued, the numbers of tables and columns in a relational database quickly grows beyond the point at which technical staff can easily manage them. Therefore, these staff take simplifying design decisions.

An example would be a Security Master table designed to hold information on both bonds and equities, instead of two tables (one for bonds and one for equities). This design immediately violates relational design theory because it is one table corresponding to two different types of thing: bond and equity. In a CFL a bond will have certain unique properties, such as Maturity Date (except for perpetual bonds), and Coupon Rate (even if it is zero). Equities too, will have their unique properties, such as the degree of ownership. Therefore in the Security Master table we will see one column for Maturity Date and another for Coupon Rate. These columns have no applicability to equities - they are only relevant to bonds. Yet, according to relational database theory, every record in the Security Master table must represent an instance of the single concept that the table is designed for, and all such instances must have identical properties. Clearly, this requirement is violated. After all, neither Maturity Date nor Coupon Rate can be populated for equities, but they must be populated for bonds.

Such violations are so common in relational databases that they typically go unnoticed. The data professionals involved reflexively try to reduce the number of tables in a database, so that each table represents a collection of concepts, rather than a single concept. An illustration of this is the resulting and widespread practice of introducing of additional columns, such as for example "SEC_INSTR_TYPE", in order to make the inherited internal distinctions match up to an externally understood classification scheme. For this reason the concepts in any CFL will often not correspond exclusively to any single data object in a database.

The need to avoid adding columns to database tables is often taken to an extreme in many situations, by generalizing designs in relational databases so that these databases are simply stores of data not organized in a way that directly corresponds to any external reality. It is up to program logic and human interpretation to provide this correspondence. A well-known example of this is the so-called "Party Model". The Party Model is a relational database design pattern that is very well known and widely used. It states that all natural persons and legal persons can be generalized into a single table (Party) and then the roles they play - e.g. Customer, Employee, Supplier, Regulator, - can be treated separately. This highly generalized design means that if a new form of party appears, e.g. Strategic Partner, it can be accommodated in the existing database with no changes.

The Party Model is very common in reference data of financial institutions, accommodating a whole variety of institutions in finance, e.g. Hedge Fund, Mutual Fund, Broker, Dealer, Client, Regulator, etc. Naturally, each of these types of institution has its own unique attributes, but these are captured in the generalized database design. This makes it very difficult to understand how the data corresponds to reality, and very often teams of data specialists have to be maintained to do this. However, firms find this approach preferable to dealing with changes to relational databases. The point is, of course, that in such environments - which are the overwhelming majority - there can be no direct correspondence between a CFL and database.

Semantic adapters as a possible means of aligning databases and a CFL

This conflict between the concepts of a common financial language and practicalities of database design outlined in the previous sub-section represent a considerable challenge, because the regulators want a CFL to promote data exchange, and for this to happen the data in any database must ultimately be made to correspond to a CFL.

The natural reaction of data professionals is to map data in any particular database to a file format - typically for regulatory reporting. These interfaces are created one at a time, and ultimately there is no real guarantee of any consistency between them. However, the vision of a CFL is much more oriented to making general requests for data, such as the US Office of Financial Research is empowered to do. For this to happen within any acceptable timeframe, a database really must have some form of correspondence to a CFL.

Figure 2 illustrates the problem and provides a possible solution.

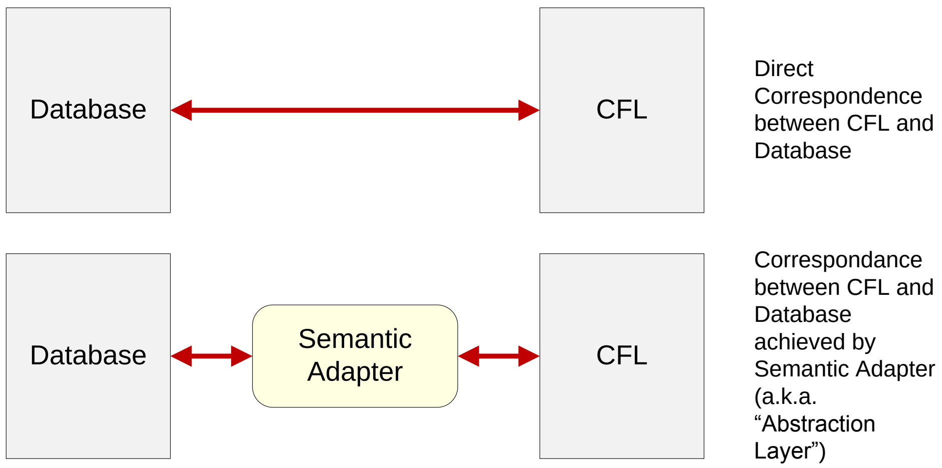


Figure 2: *Relationship of Database and CFL*

Figure 2 presents the idea of a layer between the database and the CFL to render the data in the database as a "view" of the CFL. This layer is often called a "Semantic Adapter", although sometimes it is termed an "Abstraction Layer".

This technique has in fact been attempted in large scale data environments within enterprises for a number of years (for example the global internal exchange of information within large international financial institutions). In an attempt to exchange data among applications, near-real time messages containing packets of data have been sent though infrastructural components called Enterprise Service Buses (ESB's). These messages are "listened to" by applications connected to an ESB. Each such application has a Semantic Adapter it uses to translate the data content of the message back to its underlying data store definitions.

The best practice for such data exchange is to use XML-based messages that are conformed to a single standard termed a "Canonical Data Model", which is typically an XML schema ("XSD"). In practice, however, it has proven extremely difficult for enterprises to create Canonical Data Models. Problems include:

- What is the scope of such a model? It is intended to be universal within an enterprise, but is this really for a particular business unit, or subject area, or even application? In reality, the Canonical Data Model often just turns out to be a library of message formats, each independent of the other, with no real integration, and thus no aspect of a common language. Individual programming teams typically construct message formats in isolation from others.
- Semantic issues often remain unresolved. This is because the Canonical Data Models are generalized to just become containers whose content has to be interpreted by special program logic. E.g. An element might be specified for "Instrument Type" and the users told they can "put anything they like" into the element "to represent any financial instrument". Semantic analysis is then not done well.
- Identity issues remain unresolved. The identity of any instance of anything (e.g. of a counterparty) is not addressed. Admittedly, the LEI initiative is addressing this for counterparty, but there are plenty of other types of things where instances can be specified in multiple ways leading to identity resolution conflicts.

Very few Canonical Data Models have been produced for something like an industry. The Global Justice XML Data Model (GJXDM) is one.[27] This has been produced by the US Federal Government to promote data interoperability among the many law enforcement organizations in the different jurisdictions across the USA. Its scope, however, is limited to data exchange, and is not really "common" in the sense that the CFL is. It is intended to be a reference for constructing individual message types for exchange among law enforcement organizations. This is a much more limited goal than the CFL has, and the difficulties referred to above still remain. However, the GJXDM has many technical features that might

[27] https://it.ojp.gov/jxdm/

be of interest to a CFL and is perhaps one of the very few examples outside of finance from which lessons may be learned.

The idea of a Semantic Adapter between any database and the CFL is therefore technically possible, but appears to be very difficult to achieve in practice. Furthermore, a Semantic Adapter would have to be built between every database and the CFL. This could not be got round, e.g. by creating one data store in every financial institution that faces the CFL, because all the data stores contributing to this data store would have to be conformed to it, which is essentially the same problem. The net effect would still be to require every database to create a Semantic Adapter to the CFL.

Of course, the scope of any particular database is dramatically less than the scope of the CFL (which is universal for finance) and this alone would present enormous difficulties, because the semantics of each database would have to be located amongst a much vaster CFL, which is a considerable challenge for data analysts.

Lessons from failure: the Corporate Data Model

There is a sense in which the idea of a CFL is not a new one in data management. In the latter half of the 1980's and into the early 1990's large sums of money were spent by large enterprises in all industrial sectors in attempts to create Corporate Data Models (CDMs).

The idea of CDMs came from the revolutionary success of relational database technology, combined with the later development of Entity-Relationship Modeling (ERM). Originally, relational databases were specified by hand by programmers. In 1976, Peter Chen of Louisiana State University developed a visual modeling notation - ERM - that permitted these designs to be viewed in a graphical format.[28] At last a blueprint for a database could be viewed before building it, just as a blueprint for a house can be viewed before it is built. This new technique gave rise to the discipline of Data Modeling, which had not really existed before, and meshed seamlessly with the relational theory of Codd and others.

The idea quickly developed that if the entire data "universe" of an enterprise could be put into a single data model, then this would automatically achieve data integration. Rather than individual databases being built in isolation and then "mapped" to each other

[28] (Chen, 1976)

afterwards, a single CDM would allow each database to be a subset of the overall model, thus avoiding any future mapping.

Despite the resources allocated to these efforts, the promise of the CDM was never really achieved. Reasons include:

- A data model is as much a design activity as it is an analysis one. As noted above, many design decisions are made for optimizing data storage, or speed of retrieval, or speed of development, or having fewer objects to administer. Design decisions of this kind create the very difficulties that the CDM was intended to avoid - and "mapping" then has to occur between different components. For a long time there was an argument in the Data Modeling community as to whether Data Modeling was predominantly analysis or design. Today, it has largely been resolved in favour of design. This means that any set of business data could be modeled in any number of different ways, depending on the design criteria adopted. In such a context a unified CDM is clearly impossible.

- The CDM took too long. Although resources were made available, typically only a few parts of the CDM were developed after a couple of years. The level of detail of the CDM was a fully attributed data model (i.e. all attributes defined for all entities, with all relationships specified). This outcome was in turn a result of not being able to get detailed inputs from subject matter experts (SMEs) who knew aspects of the business, or actually having gaps in knowledge about the business. The scope of the CDM project was also huge with respect to the data modeling resources available.

- Change was another factor. As soon as one section of the CDM was complete, modelers moved on to another section, but the part of the business represented by the completed section began to change, and modelers had to go back to it to keep it up to date.

- Semantic conflicts. For a CDM to work, each concept must have a single stable definition. This turned out to be impossible. For instance, the term "Customer" is very important, but there is no single definition of it. Marketing often includes prospects as Customers, whereas for Accounts Receivables, a Customer is someone who has paid money or owes money to the enterprise. Such problems led to enormous difficulties in dealing with important broadly shared concepts. Methods to solve such issues included generalizations (e.g. "A Customer is something of interest to the enterprise") or voting to force one view on everyone else, or Data Modelers simply selecting a view. Attempts to truly analyze what is meant by a concept such as Customer took a long time, and contributed to management losing patience with the entire CDM initiative.

From this it can be seen that attempts to build a single universal view of a large scope of data are unlikely to be successful.

The need for multiple models

The flawed assumption of both CFL and the unsuccessful CDM exercises of the past, is that there is a single underlying reality that is ultimately captured in data and can thus be expressed in a single representation, be it a model or language. As discussed in Section 2, there are several reasons for doubting that this is actually the case. We can explore further the example of a mortgage loan in the context of database modeling and design. Is there a single model that can represent everything about a mortgage loan, or are there different contexts in which the mortgage loan presents different semantics? That is, can the concept of "mortgage loan" be placed in different business-meaningful models where it will have different properties and participate in different relations? If this is so, then there is reason to suppose there is no single underlying reality that can be captured in data.

Table 1 provides examples of attributes and relationships of the concept Mortgage Loan, each unique to one of the different concept systems a mortgage loan or mortgage note might appear. "Concept System" is defined in ISO 26162 as "a set of concepts structured according to the relationships among them".[29] It is not as specific as "model" (since something has to be built from a model) and not as vague as "context"; instead it focuses only on concepts.

[29] ISO 26162 (2012), "Systems to manage terminology, knowledge and content — Design, implementation and maintenance of terminology management systems".

Concept System	Attribute	Relationship
Loan Origination	Appraisal Fee	Processing by Closing Agent
Loan Servicing	Loan Payment Received Date	Sale to New Servicer
Collections	Robocall (automated calling to seek debt repayment) Frequency	Skip Trace Processor (individual responsible for locating the whereabouts of an indebted customer)
Foreclosure	Recovery Value	Judicial State Attorney
Securitization Pool Selection	Date of Addition to Pool	Assignment of Note to Trust
Securitization Accounting	Collateral Servicing Fee	Document Custodian

Table 1: *Examples of Unique Attributes and Relationships for Mortgage Loan Across Different Concept Systems.*

From Table 1 it can be seen that the concept Mortgage Loan can feature in a number of different concept systems, each of which is oriented to a specific area of business. Within each concept system there are attributes that, for Mortgage Loan, are only possessed insofar as the concept appears in that concept system, and are not shared with any other concept system. Further, in each concept system, there are relationships in which Mortgage Loan is a participant which do not occur in any other concept system.

Within all of these concept systems, the definition of Mortgage Loan - its essence - remains unchanged. Therefore we have the same concept that appears and behaves differently depending on the perspective of what concept system it is being considered. For instance, it is meaningless to discuss assignment of a mortgage to a securitization trust in the Loan Origination concept system - after all no loan exists until the process involved in this concept system is complete. No data model for a loan origination system would ever include any construct for a trust.

It could be asked, why not just aggregate all the attributes and relationships of Mortgage Loan from each of these concept systems into a single data model that could then be used to specify a data store that unifies all the concept systems (akin to the vision for the Corporate Data Model).

If such an exercise were undertaken, it would be necessary to overlay the concept systems on such a data store, so workers could understand which concept system each data element and relationship pertained to. This is because each concept system carries with it a set of processing rules oriented to the overall task that it is supposed to perform.

Perhaps, however, the data about Mortgage Loan would not be used for processing, but for answering queries. In this case, queries might cross concept systems. E.g. do Mortgage Loans with high Appraisal Fees default more frequently? It would seem that we just need to have an environment where all the data is available and thus that there should be a single unified model. However, considerable experience in data management weighs against this view. Pure relational database designs have been found to be inadequate for reporting and analytics. They suffer from extremely slow performance for general reporting, and are far too complex for business analysts to easily navigate. This was the driver for the emergence of Data Warehousing in the 1990's which deliberately broke the rules of database normalization, replacing them with a different set of rules oriented to effective analysis and reporting. Additionally, Data Warehousing addressed the requirement of historical data, which pure relational theory never had.

Without going into detail about techniques such as Dimensional Modeling, Changed Data Capture, and Star Schemas, it is possible to summarize Data Warehousing as follows:

- The information requirements of the analysts - the data they need and the families of queries they want to run - are first understood.
- Data stores are shaped to answer these information requirements, typically by creating a number of 'data marts' (as defined by Wikipedia a 'data mart is a subset of the data warehouse that is usually oriented to a specific business line or team.'). Each family of queries requires a different data store.
- The data stores are then populated, with best efforts to deal with data integration and quality issues.
- Business analysts then query the data stores.

Essentially, the concept systems in Data Warehousing are dictated by the types of queries business analysts want to run. Each type of query becomes a new concept system that is turned into a data model, and then a corresponding data store is built. There is no single unified concept system or data model.

Today, even larger quantities of data are being processed in "Big Data" environments than have been common in Data Warehouses. However, the central idea of reshaping the data to look like the analysts' query is still the same.

From this discussion it can be seen that attempts to build large scale unified models in Data Management have had poor outcomes. Yet there is a need to build smaller models corresponding to distinct concept systems serving particular business functions. The same concept can legitimately appear in many such models, although it will have some attributes and relationships that are unique in each.

This corresponds to the view taken by the terminological community in the ISO 26162 standard (the ISO standard for the design, implementation and maintenance of terminology management systems). Terminologists, in addition to data modelers and ontologists, are a community that is interested in understanding what concepts are and how they interrelate. A practical rule for them is that a concept system should contain no more than 200 concepts. This keeps the scope of a concept system to something that is manageable and achievable in a reasonable timeframe, and which is directed towards some well-understood business goal.

This suggests that CFL applicable across the entirety of wholesale financial services will be no more than a limited number of common concepts, usable in many different concept systems (these would for example include general accounting concepts such as 'asset' and 'liability'). Most practical CFL will be part of its own concept system in which unique attributes and relationships are fully specified, so that it can be oriented to serving a particular business function and 'community of interest'.

4. Lessons from case studies

This section discusses three case studies of common language and draws lessons from these case studies on the benefits that can be obtained from common financial language. These three case studies are FIX, the messaging standard for pre- and post-trade messaging in financial markets; ISO 20022, the meta standard for creation of financial messages; and the global LEI system.

Case study 1: the FIX protocol

One of the most successful common financial languages employed in financial services is the FIX protocol.[30] FIX was launched in 1992-1993 as a joint venture between the US broker-dealer Salmon Brothers and Fidelity investments.[31] FIX was an immediate success and in the years following the January 1997 launch of the FIX protocol version 4.0 became the almost universal standard for pre-trade and trade-execution messaging in cash equity markets.[19]

An advantage of choosing FIX as a case study is that it is a public domain standard for pre-trade communication with its language fully documented on the FIX website. The documentation for version 4.0 (FixProtocol, 1997) explains the structure of FIX messages. A FIX session is defined as a bi-directional stream of ordered messages between two parties within a continuous sequence number series. A single FIX session can exist across multiple physical connections; parties can connect and disconnect multiple times while maintaining a single FIX session. Connecting parties must bi-laterally agree as to when sessions are to be started / stopped based upon individual system and time zone requirements. It is recommended that a new FIX session be established once within each 24 hour period.

Each message, administrative or application, is preceded by a standard header (contains 7 required tagged fields, including the sequence number, and 12 optional fields) and is terminated by a standard trailer (containing 1 required field – the checksum – and 2 optional fields). The header is used to identify the message type, length, destination, sequence number, origination point and time; and may communicate other information about participants to the trade. The trailer is used to segregate messages and contains the three digit character representation of the Checksum value.

There are two categories of message

- Administrative messages, used for operation of the link, include logon , "heartbeat" that checks connection is still live and Test Request for heartbeat, resend request,

[30] For more detailed review and assessment see (OXERA, 2009).
[31] See (FixProtocol, 2001)

reject message (allowing for text description of reason for reject), sequence reset, and logout

- Application messages. There are 18 of these in FIX 4.0:

1. Advertisement, announced completed transactions
2. Indication of interest (used by brokers)
3. News, a general free format message format
4. Email, similar to news but private for two parties
5. Quote request, to reflect practice in some markets
6. Quote (can be used to respond to quote request or sent unsolicited)
7. New order – Single, used by institutions for security and forex orders
8. Execution report, used for various tasks, confirming order receipt, confirming cancel or replace requests, relying order status, relaying fill information as orders are worked, rejecting orders, reporting fee calculations. NB there are *thirteen possible order statuses*, from New/ Pending New to Filled/ Done for the Day, with various intermediate or cancelled states and a post trade “calculated” status.
9. Don’t Know Trade – a special message used in one-way execution reporting, a form of execution rejection message, identifying reasons
10. Order Cancel/ Replace, only accepted if the order can be pulled back, applies to entire order
11. Order Cancel Request, used to cancel the remainder of an order
12. Order Cancel Reject message (used if 10. Or 11. cannot be accepted)
13. Order Status Request
14. Allocation, instructs broker how to allocate executed shares to sub-accounts
15. Allocation ASK, used by broker to acknowledge receipt and status of Allocation message
16. New Order List, provides a group of orders (instead of a single order under 7.) but is not executed until receipt of List-Execute message, when it is executed in specified order. Execution is not normally reported order by order (unless specifically requested).
17. List Execute/ List Cancel Request
18. List Status Request/ List Status. Request for information on orders in list (limited to 50 orders per message) and response.

These various FIX 4.0 messages make use of 140 defined fields (with one field, always the third field in the message, defining the message type). These fields and the acceptable entries within them constitute the grammar and vocabulary of FIX 4.0 messaging.

This grammar and vocabulary expanded enormously, through successive versions of FIX (4.1-4.4), until by the December 2006 release of FIX5.0, there were 101 types of application message and 1139 defined message fields.[32] This expansion supported the introduction of a wide range of new functionality for messaging in fixed income, money market, foreign exchange, cross-border and derivative transactions and for post-trade processing; and meeting the needs of a much wider range of entities, not just broker-dealers and their customers, to use FIX.

At the same time the much greater complexity and functionality has made FIX5.0 a more difficult language to learn, with greater possibility that implementation does not fully adhere to the standard. Consequently, as FIX has grown, it has supported an entire 'ecosystem' of service providers who offer firms FIX implementations that can be integrated into their own business processes.

FIX has been an undoubted success. It illustrates very well how standardisation can support both automation (modern computerised trading would be very difficult to conduct as it is today without the FIX standard) and the consolidation benefits of being able to replace multiple communication channels to different brokers with a single standardised interface.

FIX though also illustrates the difficulties of getting standardisation accepted even where it is of undoubted customer benefit. FIX, while now almost universally employed in pre-trade and execution messaging in equity markets, has made much less headway in fixed income and derivatives markets, or in post-trade messaging. This is partly because of the challenge of achieving critical mass of users. Like any other communication standard the main benefits of FIX are really only a large proportion of other users have also adopted the same standard. This lack of uptake appears to be because of the perception that the FIX standardisation has supported a much greater degree of competition in equity brokerage with an accompanying

[32] Source: FIX 5.0 documentation. FIX 5.0 also introduced a 'transport' independent version of FIX, separating the administrative messages as a separate transport standard (FIXT) and so allowing the application messages to be sent using any chosen communication protocol.

sharp fall of broker margins.[33] With this experience in mind, fixed income and derivative brokers are relatively cautious about adoption of FIX compared to its original users in equity trading.

Case study 2: ISO 20022.[34]

There are many other communication standards widely used across the financial services industry aside from FIX. This second case study focuses not on any individual financial messaging communication standard, but instead on the overarching *process* standard for financial messaging, ISO 20022, developed by SWIFT and ISO (International Organization for Standardization).

As described in (SWIFT, 2010) ISO 20022 is a 'recipe for making financial messaging standards', setting out in a structured way a common approach to establishing and then implementing standards for financial messaging. ISO 20022 can thus be described as a 'meta-standard' created in order to overcome the challenge of sometimes substantial differences in both syntax (structure) and semantics (meaning of individual message components) in different financial services messages. Existing alternative and sometimes competing messaging systems include ISO15022 for cross border securities settlement, ISO8583 for credit and debit card transactions, FIX, FpML for financial derivative transactions, as well as SWIFTs own proprietary MT messages used on the SWIFT network for correspondent banking, foreign exchange and trade finance.

The approach offered by ISO 20022 is in many respects similar to that used in the development of business ontologies for computer systems and in information management. Like these ontologies ISO 20022 distinguishes the business concepts from the representation in computer readable form. ISO 20022 consists of three separate 'layers':

- Identification of key business processes and concepts
- Logical (communication standard independent) message construction
- Implementation in a particular communication standard ('syntax')

[33] The causes of the major decline in equity brokerage margins over the past 30 years are still an open research topic. While message standardisation may well have played a role, it is certainly not the only factor, others include the shift to automated execution via order-driven trading platforms; the development of algorithmic and high frequency trading; regulatory and customer insistence on demonstrated 'best execution'.
[34] This section draws on (SWIFT, 2010)

Using ISO 20022 allows systematic identification of the various elements of a financial process and the parties involved in it (for example in the context of a payment, these might be the paying party, the receiving party, the bank of the paying party, the bank of the receiving party, and the payment itself.) All these elements and parties are represented diagrammatically, using Object Management Group's Unified Modelling Language (UML).

This in turn allows identification of all the message elements that are required in order to execute the process and how they relate to each other – the logical message construction. Finally the message itself can then be formulated in a particular context – today only XML and ASN.1 syntaxes are recognised for ISO 20022, but others are possible. ISO 20022 incorporates a set of design rules to map messaging to XML based messaging syntax. Using XML facilitates machine readability and provides the easiest mapping of the message to different formats and standards. However, the design of the standard intentionally de-emphasizes syntax, because the syntax you use depends on many technical and historical factors

The functionality of ISO 20022 is delivered by its repository – a web-based dictionary that defines all the components of both business processes and of financial messages used in the various financial transactions covered by the standard *and* provides links to tables of all the information that may be needed to refer to these components in the construction and execution of a financial message. Using this repository provides a method for developing financial message definitions that are consistent in terms of structure meaning both internally and with other ISO 20022 messages. According to (SWIFT, 2010) this repository 'holds several hundred business components, around 700 message components and more than 250 message definitions'. ISO 20022 allows firms to develop payments and other financial messaging capacity in their own business systems at low cost and with minimal problems of compatibility with the systems used in other firms.

While the advantages of ISO 20022 are evident, its take up and usage in different business areas from its initial launch in 2004 was at first rather slow. This is not surprising. In contexts where messaging arrangements are already implemented and largely satisfactory, then there is relatively little benefit to a business to commit itself to a meta standard for the development of new messaging facilities. It is likely to be lower cost to simply manage with existing cumbersome updating processes rather than have staff learn an entirely new approach. Over time though, ISO 20022 has become much more widely used, largely in the context of major new projects and initiatives. So for example ISO 20022 is the standard used for exchanging payment details in the Single European Payments Area (SEPA) scheme. It has also been selected for the messages used by the ECB's forthcoming TARGET2-Securities settlement system, and will be adopted by Europe's high value payments system, TARGET2.

Over time further migration to using ISO 20022 can be expected, in particular where businesses want to obtain the advantages of XML messaging but also wish to maintain backwards compatibility to older messaging standards where these continue to be used by some customers or counterparties, or where they wish to maintain an old standard in some internal legacy system. The translation between different message syntax supported by ISO 20022 is well suited to maintaining such 'bilingual' messaging arrangements. In this way ISO 20022 is a semantic adapter, very similar to that described in Section 3 above in the context of data storage and retrieval.

While ISO 20022 successfully supports interoperability of different financial messaging systems, like FIX its range of application remains limited to financial transactions. It cannot be applied to collecting and summarising information held in different databases. We turn next to an example which promises to be especially valuable in this latter context.

Case study 3: the global LEI system.[35]

In June 2012 the Financial Stability Board announced that it was to establish a global system for the unique and unambiguous identification of counterparties and clients in wholesale financial markets. This was a global extension of the 2010 Dodd-Frank act and the European Market Infrastructure Regulation (EMIR) which required a similar identification schema for participants in US and EU markets, when it became apparent that such a scheme was only fully workable if implemented at a global level.

To date the main application of LEI has been in the implementation of the Dodd-Frank requirements for recording of OTC derivative contracts in trade repositories. The CFTC – the US regulator responsible for implementation of this part of Dodd-Frank - has commissioned DTCC together with SWIFT to provide so called interim compliant identifiers (CICIs). These

[35] This sub-section draws on Chan and Milne (2013).

are one example of so called 'pre-LEIs' identifiers (others are being issued in Germany, France and the UK and will soon be issued in other jurisdictions) that comply with the requirements announced so far for the issue of eligible LEIs. The full LEI system will not be launched until international governance arrangements are in place, with a central operating unit (to be located in Switzerland) operated by a charitable foundation, the LEI foundation; and oversight by a group of international regulators (the Regulatory Oversight Committee). These arrangements will ensure that the LEI's issued by various local operating units in different jurisdictions comply with globally set standards for data quality and verification, and accepted as identifiers across the world.

The potential benefits from the LEI system are very large, both in terms of improving efficiency (reducing costs) at financial institutions and supporting better quality data and information for risk management and regulatory oversight. At present there is no standard identification system for use across the industry in customer onboarding, in maintaining KYC (know your customer) and AML (anti-money laundering) systems, or for supporting straight through (automated) transaction processing. Potential cost savings to the industry from removing duplication and reducing manual interventions in these systems look capable of removing as much as $10bn per year or more of current costs from the financial institutions participating in wholesale markets.[36]

They will also support the growth and development of wholesale financial intermediation in emerging markets, where identification of counterparties can be relatively difficult. Legal entity identification is a persuasive example of the potential benefits of clear and unambiguous identification (further cost savings, although perhaps on a smaller scale, could also be obtained from similar standardized identification of financial instruments).

This is not to say that achieving these benefits will be quick or easy. It will take some time, a matter of years, before these they are fully realized. This will require that the LEI is adopted as the standard counterparty identifier in a wide range of security, derivative and credit markets; and in most if not all financially open jurisdictions. Internal systems for many business processes and reference data management may need re-engineering to take full advantage of LEI. But the business case for adoption seems to be a strong one.

[36] For discussion of the source and magnitude of cost savings from the LEI see (Chan & Milne, 2013)

Lessons from these case studies

These case studies illustrate three sources of benefits from common financial language. FIX illustrates the benefits of *standardization*: ensuring that information is recorded and represented and understood in the same way. This is a prerequisite for improving the efficiency of business operations, through automation (removing the need for manual interventions) and consolidation (reducing fixed costs by reducing the number of platforms and communication channels that must be maintained).

LEI illustrates another source of benefits, that of unambiguous *referencing*. Unique and unambiguous identification of counterparties and of financial instruments are the key to eliminating the large amount of duplication of business processing i.e. ensuring that the same job is not undertaken twice.

While there are clear business benefits to both standardization and to unique identification, our case studies illustrate that fragmentation of both standards and identifiers may persist for considerable periods. The spread of FIX beyond its initial user base in equity market trading has been limited. It has achieved such widespread usage in equity markets that any participant who wishes to participate in these markets is pretty much obliged to use FIX. But uptake has been much more limited in fixed income, derivatives and foreign exchange. In the case of LEI it appears that the continued fragmentation of identifiers has been a co-ordination problem. It has taken regulatory initiative to overcome the co-ordination barriers to establishing a global identifier system.

Our case studies on LEI and ISO 20022 also indicate that many if not all of the benefits that Haldane anticipates from common financial language in (Ali et al., 2012) arise from a third source of benefits: the *harmonization* of key elements of different languages. This comes into play on those occasions when different systems have to be bridged or operated together; for example for the aggregation of information at firm, market and global level.

Taking first the example of the LEI. Unique and unambiguous identification can certainly lead to substantial cost savings within a single business process (for example customer onboarding, or the execution of customer orders for a particular type of transaction). Unique identification, however, does not offer the same kind of efficiency benefits in any business task that requires combining data from different sources and from different systems.

Yes, there may be identifiers common to different systems, but this does not guarantee that the data associated with these different identifiers can be automatically and meaningfully aggregated together to provide an accurate cross-system information; or that the various financial obligations captured in these systems can be automatically divided into constituent parts, as required in a legal process such as a bankruptcy or a divestiture.

Our case study on ISO 20022 provides a further illustration of the importance of *harmonization*. The main benefit to most users of ISO 20022 is not the XML standardization of financial messaging that it supports. Most users, and their customers and counterparties, are still wedded to many older messaging syntaxes, which continue to be widely used in legacy computer systems. As a result they are not moving to full adoption of XML messaging any time soon. What they value from ISO 20022 is that enables a common approach to supporting both pre-XML legacy payment messaging and more recent XML messaging structures.

This parallels our discussion of 'Semantic Adapters' in Section 3, which illustrates why, in order to obtain meaningful aggregation (for example in risk monitoring and management) or legally valid partition (to resolve an insolvency or conduct a corporate action such as divestiture which cuts across business systems), it is typically necessary to maintain separate semantic adapters for each business system, interpreting the various elements in the corresponding database or databases in an appropriate fashion. The greater the degree of *harmonization* of terms between the different systems then the easier it will be to develop and maintain such semantic adapters.

Achieving such harmonization in a way that is useful to management and to regulators is not just a matter of 'nouns' and 'verbs' (e.g. accurate identification of both counterparties and instruments). An on-going effort will be needed to maintain appropriate semantic adapters (i.e. if you like, language translation) from systems that record the underlying granular data, to ensure that these can be used to address the questions and queries that are required for risk reporting and the monitoring of systemic financial risk; and that the same or at least similar questions can be addressed using different systems within the same firm, and for different firms.

The FIBO project is a promising effort at harmonization of concepts and terms for a wide range of financial contracts and instruments. Adoption of FIBO standards will lead to firms across the industry using similar approaches to data categorization and storage. But complete alignment of data and business systems is not a practical possibility, rather the extent of this alignment will always be a matter of degree.

5. Conclusions: a journey not a destination.

We conclude with a brief summary of the high level points that have emerged from our analysis of the proposal for a common financial language. The proposal itself is very attractive. Lack of understanding of risk exposures, both of the exposures of individual firms and of the network of exposures across the financial system as a whole, was a major problem that both contributed to the build-up of systemic risk prior to the crisis and handicapped policy makers in responding to the crisis

As argued by Andrew Haldane (Ali et al., 2012) agreement on standard identifiers (both legal entity identifiers and instrument identifiers) is an essential step to both improving the efficiency of business processes (much as the barcode and related identification standards have supported efficiency in supply chain, retailing and healthcare); and could provide a common basis for standardised communication of risk exposures (much as the development of the world wide web has facilitated business and personal communication).

Industry has and continues to make efforts to develop common language. FIX and ISO 20022 both provide examples of successful common language developed by the industry. A more far reaching initiative is the FIBO ontology which aims to provide a conceptual framework and a mapping to computer readable formats (in the OWL language) covering most aspects of financial services and thereby providing an opportunity for further improvements in both operational efficiency and the measurement and management of risk.

While we share the aims of both Haldane and of FIBO, we conclude that common language should be seen as an aspiration, a goal which industry and regulators will continually strive towards. It is not an achievable and measureable objective that can be achieved within any foreseeable timeframe. There are several reasons for this. As we discuss in Section 2, financial reality is complex and evolving, and so cannot be fully captured by any agreement on language. Even if there were a stable underlying financial reality, the language used to describe it will itself be complex and fluid. It should come as no surprise, at least to anyone with even a casual acquaintance with the philosophy of language, that it will never be possible to create a fully defined single common financial language suitable for all purposes and applications.

This does not mean that there are not substantial benefits to be had from developing more common financial language for communication between computer systems and for interaction with the databases where financial information is stored. Such standardisation can eliminate unnecessary duplications of business activities and communications; and reduce costly manual interventions in business processes. It can also help improve the quality and timeliness of risk reporting.

While more common financial language for use by computers is clearly worth having, our discussion in Section 3 suggests that a single common financial language for communication between computer systems is not a realistic practical proposition. The constraints of database design and management mean that there will always be inconsistencies between the financial reality which a ‘common’ financial language is meant to describe and the structure and organisation of underlying data. So even if there were a single unchanging financial reality (which there is not), it would not be practically possible to employ a single protocol or syntax for all possible requirements for computer processing and data storage in financial services.

Instead it will be necessary to maintain translations (or in database jargon Semantic Adapters) which translate the questions and queries asked by risk-managers and regulators or required when resolving an insolvency, into a form in which they can meaningfully be answered by any particular database. This is similar to the situation in financial messaging where ISO 20022 can serve as a mechanism for translating between different legacy financial messaging formats.

This does not mean that greater commonality of language is not worth having. Our examination of the practical benefits of more common language in our various case studies

(Section 4) has highlighted the benefits available from *standardised communication* (FIX) from unique and unambiguous *identification* (LEI) and from developing *meta standards* and *ontologies* (ISO 20022 is a successful example) which capture a shared understanding of both business processes and the terms and communications used in the systems that support them.

What seems to be needed to facilitate the collection and aggregation risk exposures and other data, both within individual firms and for the financial system as a whole, is greater *harmonization* of concepts and terminology. This in turn will make it relatively easy to create and maintain the necessary ‘semantic adapters’, so that when different systems have to be accessed to obtain and summarise information, this can be done as effectively and quickly as possible. This task requires the development of common financial language on a case by case basis, in the context of particular processes and ‘communities of users’. Development of common financial language is thus more of a journey than a destination.

Appendix: An Investigation of the functionality of the Campbell R Harvey Hypertextual Financial Glossary

This appendix summarises an effort carried out by the co-authors to investigate the functionality of the most widely used reference glossary in Finance, the Campbell R Harvey Hypertextual Financial Glossary.[37] This is not the only available glossary of this kind. Wikipedia and Investopedia both provide definitions of many financial terms and often give more extended examples of the application of particular terms than the Campbell R Harvey glossary. The Campbell R Harvey glossary is however incorporated as part of the data services provided from Bloomberg terminals and therefore has some claim to being the most important such glossary.

Our objective in investigating the functionality of the glossary was to find out if it could be regarded, even approximately, as being a dictionary for a common financial language. We therefore distinguished a number of different functions ('use cases') that we might expect a dictionary of a common financial language to provide and examined whether each of these uses was supported by the Campbell R Harvey glossary.

Find a definition for a known term in English.

This use case is completely satisfied by the glossary. However, the user must know the term to begin with. Since many terms are technical, knowing the term is likely to be accompanied by some knowledge of its definition. This use case does not help understand why a user would want to find a definition for a term they know (e.g. it could be to be informed or to be reminded).

Find all terms in which a given word or phrase is mentioned.

This use case is completely unsatisfied by the glossary. No search functionality of any kind is available.

Find all definitions in which a given term is mentioned.

This use case is completely unsatisfied by the glossary. No search functionality of any kind is available.

[37] The glossary can be accessed via http://people.duke.edu/~charvey/Classes/wpg/glossary.htm

Determine if a term is a specific concept (species) of a more general concept (genus).

(a) Tried with "Interest Rate Swap". The glossary says "A binding agreement between counterparties to exchange periodic interest payments on some predetermined dollar principal, which is called the notional principal amount. For example, one party will pay fixed and receive variable." It does not identify "Interest Rate Swap" as a type of any more general financial concept (it is defined simply one form of binding agreement) and certainly not as a type of swap.

(b) Tried with "Interest Rate Risk". The glossary says "The chance that a security's value will change due to a change in interest rates. For example, a bond's price drops as interest rates rise. For a depository institution, also called funding risk: The risk that spread income will suffer because of a change in interest rates. Thus it does not indicate "Interest Rate Risk" is a type, or class, of risk.

(c) Tried with "Market Risk". The glossary says "Risk that cannot be diversified away. Related: Systematic risk". Thus it does indicate (albeit not explicitly) that "Market Risk" is a species of risk.

(d) Tried with "Systemic Risk". The glossary says "Risk common to a particular sector or country. Often refers to a risk resulting from a particular "system" that is in place, such as the regulator framework for monitoring of financial institutions." Thus it does indicate (albeit not explicitly) that "Systemic Risk" is a species of risk.

From these observations it can be concluded that the glossary only inconsistently satisfies this use case. Furthermore, the glossary does not have a formal way to identify that a concept is (or is not) a species. Therefore, the user has to infer this attribute from reading the definition.

Identify the generic concept (genus) to which a specific concept (species) belongs.

The same experiments as with the previous use case lead to similar failure. From these observations it can be concluded that the glossary only inconsistently satisfies this use case. Furthermore, the glossary does not have a formal way to identify what genus (or genera) a species is subordinate to. The user has to infer it from the definition.

Determine if a term is a part of a whole.

(a) Tried with "USA". The glossary says "The three-character ISO 3166 country code for UNITED STATES". It does not identify that "USA" or "United States" is a country that makes up the world. This could be held to be common knowledge and so not required of a glossary of finance.

(b) Tried with "Long Leg". The glossary says " The part of an option spread in which an agreement to buy the underlying security is made." It does indicate that "Long Leg" is a part of an Option Spread.

(c) Tried with "Back Office". The glossary says "Brokerage house clerical operations that support, but do not include, the trading of stocks and other securities. All written confirmation and settlement of trades, record keeping, and regulatory compliance happen in the back office." This does not clearly indicate that the "Back Office" is a part of an enterprise. The definition is also incorrect in stating that "Back Office" is confined to Brokerage Houses. Finally it gives no indication of the extent to which records are stored in computerised databases or of the role of technology in back office operations.

(d) Tried with "Front Office". The glossary says "Refers to revenue generating sales personnel in a brokerage, insurance or other financial services operation." This is even less clear than "Back Office" in indicating that "Front Office" is a part of an enterprise. From these observations it can be concluded that the glossary only inconsistently satisfies this use case. Furthermore, the glossary does not have a formal way to identify that a concept is (or is not) a part of some whole. Therefore, the user has to infer this attribute from reading the definition.

Identify the whole to which a part belongs.

Applied the same experiments as in the previous use case. From these observations it can be concluded that the glossary only inconsistently satisfies this use case. Furthermore, the glossary does not have a formal way to identify the whole to which the part belongs. Therefore, the user has to infer this attribute from reading the definition.

Determine if a term is a generic concept (genus) which comprises a number of specific concepts (species).

(a) Tried with "Swap". The glossary says "An arrangement in which two entities lend to each other on different terms, e.g., in different currencies, and/or at different interest rates, fixed or floating.". It does not indicate that there are any types of swap, and that "Swap" is a generic concept. This information is not just missing in terms of there being no explicit statement - the user cannot infer it.

(b) Tried with "Mortgage". The glossary says "A loan secured by the collateral of some specified real estate property which obliges the borrower to make a predetermined series of payments.". It does not indicate that there are any types of mortgage, and that "Mortgage" is a generic concept. This information is not just missing in terms of there being no explicit statement - the user cannot infer it. E.g. one taxonomy would be "Forward Mortgage"; "Reverse Mortgage". "Forward Mortgage" is not in the glossary, but "Reverse Mortgage" is. Another taxonomy might be "Fixed Rate Mortgage"; "Adjustable Rate Mortgage". "Adjustable-rate Mortgage" is in the glossary, but "Fixed Rate Mortgage" is not.

(c) Tried with "Risk". The glossary says "Often defined as the standard deviation of the return on total investment. Degree of uncertainty of return on an asset. In context of asset pricing theory. See: Systematic risk.". It does indicate that there are any types of risk, and that "Risk" is a generic concept. This is not stated explicitly. However, from the mention of Systematic Risk the user might be able to infer that there are different types of risk - but this is uncertain. There are of course different type of risk - "Interest Rate Risk", "Counterparty Risk", "Systemic Risk', and so on.

(d) Tried with "Bankruptcy Code". The glossary says "Laws governing bankruptcy proceedings for corporations, municipalities, and individuals. Enacted through the U.S. Federal Bankruptcy Reform Act of 1978." It does not indicate that Bankruptcy Code is a generic concept and that there are any types of Bankruptcy Code, e.g. "Chapter 7"; "Chapter 9"; "Chapter 11"; and "Chapter 13". This information is not just missing in terms of there being no explicit statement - the user cannot infer it.

From these observations it can be concluded that the glossary does not identify that a term is a generic concept (genus) with a few possible exceptions.

Identify all the specific concepts (species) that are subordinate to a generic concept (genus).

(a) Tried with "Swap". None of the different kinds of swaps are identified.

(b) Tried with "Mortgage". None of the different kinds of mortgages are identified.

(c) Tried with "Risk". Only Systematic Risk is identified as a type of risk. None of the other types of risk are identified.

(d) Tried with "Bankruptcy Code". No genus is identified.

From these observations it can be concluded that the glossary only very rarely identifies any specific concepts subordinate to a generic concept.

Identify all the specific concepts (species) that are co-ordinate to a particular specific concept (species).

This is a combination of 2 use cases: (a) Identify the generic concept (genus) to which a specific concept (species) belongs. This use case is substantially unsatisfied in the glossary. (b) Identify all the specific concepts (species) that are subordinate to a generic concept (genus). This use case is substantially unsatisfied in the glossary.

Identify all parts of a whole.

(a) Tried with "Treasury". The glossary says "US Department of the Treasury, which issues all Treasury bonds, notes, and bills as well as overseeing agencies. Also, the department within a corporation that oversees its financial operations including the issuance of new shares." However, throughout the glossary there are entries for departments of the US Department of the Treasury. These include: "Office of Financial Research"; Office of Thrift Supervision"; "United States Customs Service". Therefore, in this case the parts of the whole are not identified.

(b) Tried with "Bond". The glossary says "Bonds are debt and are issued for a period of more than one year. The US government, local governments, water districts, companies and many other types of institutions sell bonds. When an investor buys bonds, he or she is lending money. The seller of the bond agrees to repay the principal amount of the loan at a specified time. Interest-bearing bonds pay interest periodically." Bonds do have components, e.g. a Principal Amount; Coupon Payments. Principal Amount is referenced in the definition above (albeit not hyperlinked). Coupon Payments are not mentioned in the definition but are in the glossary. Thus there is some identification of parts of the whole.

(c) Tried with "Option". The glossary says "Gives the buyer the right, but not the obligation, to buy or sell an asset at a set price on or before a given date. Investors, not companies, issue options. Buyers of call options bet that a stock will be worth more than the price set by the option (the strike price), plus the price they pay for the option itself. Buyers of put options bet that the stock's price will drop below the price set by the option. An option is part of a class of securities called derivatives, which means these securities derive their value from the worth of an underlying investment." Strike Price is identified; Underlying Security is indirectly identified; Premium is not really identified. Thus, there is some identification of parts of the whole. However, these have to be filtered out from other terms in the definition, which requires knowledge.

From these observations it can be concluded: (a) A whole rarely has its parts explicitly listed as parts (b) The parts of many wholes are not listed at all for the whole.

Understand what type of thing a term is. The user has a term but does not know its definition. Further, the user does not know what type of thing it is - e.g. a class, or an instance. Unfortunately, there is no standard meta-ontology of terminology. For the purpose of this use case we will distinguish (a) a concept (b) an instance; (c) a synonym (or sign) of another term.

(a) Tried with "USA". The glossary says "The three-character ISO 3166 country code for UNITED STATES." This indicates indirectly that the term is a type of synonym - a code. However, there is no entry for UNITED STATES. Further, there is no entry for ISO 3166-1 Alpha-2 Country Code. Therefore, the base term corresponding to this synonym. It might be held that "UNITED STATES" is a common term and does not need to be in the glossary. However, "ISO 3166-1 Alpha-2 Country Code" is a specialized term and should be there.

(b) Tried with "New York Stock Exchange". The glossary says " Also known as the Big Board or the Exchange." This does not indicate that it is an instance of a concept. Rather, it treats it as a synonym. This is a technical term, and should be explained.

(c) Tried with "Exchange". The glossary says "A marketplace in which shares, options and futures on stocks, bonds, commodities, and indexes are traded. Principal U.S. stock exchanges are: New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and National Association of Securities Dealers Automatic Quotation System (Nasdaq)." This does not explicitly state that "Exchange" is a concept, but the reader can infer it from the definition.

This use case is not satisfied. Although it is an important use case, it seems to be rarely satisfied in any glossary.

The user can easily find the glossary entry for any technical term in a definition in the glossary.

(a) Tried with "USA". The glossary says "The three-character ISO 3166 country code for UNITED STATES." "ISO 3166 Country Code" is a technical term. However, it is not hyperlinked. "ISO" is hyperlinked, but it is less of a technical term than "ISO 3166 Country Code".

(b) Tried with "Interest Rate Swap". The glossary says "A binding agreement between counterparties to exchange periodic interest payments on some predetermined dollar principal, which is called the notional principal amount. For example, one party will pay fixed and receive variable." Here, all technical terms are hyperlinked, which (a) indicates they have entries in the glossary; and (b) the entries have definitions that can be read.

(c) Tried with "Certificateless Municipals". The glossary says "Municipal bonds with one certificate which is valid for the entire issue, and having no individual certificates, easing transactions. See: Book-entry securities." Here, all technical terms are hyperlinked, which (a) indicates they have entries in the glossary; and (b) the entries have definitions that can be read. This use case is generally satisfied, but not always. A few hyperlinks were checked. Nearly all (but not all) worked.

We can conclude by summarising the findings obtained from this investigation. The use cases (functionalities) examined here are some hypothetical applications of the glossary as a dictionary of a common financial language. It is by no means certain that the use cases captured here represent the majority of use cases, or even the most important that might be required of a dictionary of a common financial language. This could only be judged if the glossary were used in practice as a source of common financial language.

Still, of the 13 use cases examined, 2 were generally or completely satisfied, 4 were sometimes satisfied, and 7 were satisfied rarely or never. It appears that this very commonly used financial glossary only satisfies a relatively narrow range of use cases. There are many use cases which might reasonably be expected for a glossary of a Common Financial Language that are not satisfied at all, or materially unsatisfied.

Any Common Financial Language will require documentation in the form of a dictionary or glossary. The use cases detailed here are the kind of applications that should form a basis for designing the functionality of such a glossary. It seems that, quite apart from the conceptual objections to a common financial language reviewed in our main text, there is quite a long way to go in developing a fully functional dictionary that can support a common financial language

References

Ali, R., Haldane, A. G., & Nahai-Williamson, P. (2012). Towards a Common Financial Language, text of speech to Securities Industry and Financial Markets Association, NY. Bank of England.

Blind, K. (2004). *The Economics Of Standards: Theory, Evidence, Policy* (p. 368). Edward Elgar Publishing.

CEIOPS. (2009). *CEIOPS's Advice for Level 2 Implementing Measures on Solvency II: Technical Provisions – Article 86 f Standards for Data Quality (former CP 43)*.

Chan, K. K., & Milne, A. (2013). The Global Legal Entity Identifier System: Will It Deliver?

Chen, P. (1976). The Entity-Relationship Model: Toward a Unified View of Data. *ACM Transactions on Database Systems*, *1*, 9–36.

Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, *13*(6), 377–387.

Crivelli, P. (2008). Plato's Philosophy of Language. In *The Oxford Handbook of Plato*. Oxford University Press. doi:10.1093/oxfordhb/9780195182903.003.0009

Deloitte. (2008). *South African Securitisation Industry: Top 10 Issues for 2008*.

European Commission. (2013). *Financial Terminology Study* (p. 17).

FixProtocol. (1997). FINANCIAL INFORMATION EXCHANGE PROTOCOL (FIX) Version 4.0.

FixProtocol. (2001). Fix History as of 2001. Retrieved from www.fixprotocol.org/documents/647/fix_history_20010221.ppt

Hayek, F. A. (1980). *The Counter-revolution of Science: Studies on the Abuse of Reason* (p. 415). Liberty Fund Inc.,U.S.

Hepp, M. (2008). Ontologies: State of the Art, Business Potential and Grand Challenges. In M Hepp, P. Leenheer, A. de Moor, & Y. Sure (Eds.), *Ontology Management - Semantic Web, Semantic Web Services, and Business Applications*. Springer.

ISO. (2010). *International Standards and Private Standards*.

OXERA. (2009). *What are the benefits of the FIX protocol?* (p. 68). Oxford, UK.

Smith, R. (2000). Aristotle's Logic.

Sowa, J. F. (2000). *Knowledge Representation: Logical, Philosophical and Computational Foundations* (p. 594+xiv). Brooks Cole Publishing Co., Pacific Grove, CA.

Staab, S., & Studer, R. (Eds.). (2009). *Handbook on Ontologies* (p. 811). Springer.

Swann, G. M. P. (2000). *The Economics of Standardization* (p. 65).

Swann, G. M. P. (2010). *The Economics of Standardization: An Update* (p. 78).

SWIFT. (2010). *ISO 20022 For Dummies* (p. 52). John Wiley & Sons, Ltd.

Understanding the Legal Foundations of Business Entity Types, Ownership and Control. (2012). Enterprise Management Data Council: Financial Industry Business Entology (FIBO).

Wittgenstein, L. (2009). *Philosophical Investigations* (p. 564). Wiley-Blackwell.

Wittgenstein, L., & Russell, B. (n.d.). *Tractatus Logico-Philosophicus*. Cosimo Classics.

Wolf, M. P. (2009). Language, Philosophy of. *Internet Encyclopedia of Philosophy*.