This special issue of the *Journal of Financial Stability* collects papers on the theme of “Data Standards, Information and Financial Stability”. Many of these papers were presented at a conference of the same title held at Loughborough University, UK, on April 11th-12th, 2014. The conference was part financed by Alfred P. Sloan grant number 2012-5-50 for research on the use of a global legal identifier system to enhance financial stability. Thanks are due both to *Journal of Financial Stability* editor, Iftekhar Hasan, for welcoming the idea of a special issue on this theme, and to journal manager, Clayton Lamar, for his highly professional support throughout the lengthy process of reviewing papers and assembling them for publication.

By way of motivation for this special issue, it is worth providing a couple of illustrations of the importance of data standards and information to financial stability.

The first illustration are the data and information problems surrounding the resolution of Lehman Brothers, following its failure in September of 2008. This has been a daunting challenge, both because of the complex structure of Lehman Brothers and the insufficiency of data in the form needed for resolution of claims. The group had around eight thousand operating subsidiaries, each undertaking financial investments and business operations and raising short term finance, and all ultimately owned by a single holding company, Lehman Brothers Holdings Inc., which issued publicly traded equity and long term debt (Kapur, 2015). There was no routinely maintained and systematically organised record of Lehman assets and derivative obligations and of all the competing claims on the firm. Instead it was necessary to employ thousands of professional staff to go through a mass of disaggregated records in order to sort out who owed what to whom. The realisation of Lehman assets and positions for cash and the allocation of proceeds amongst the many claimants is a task of many years that still continues today.

One unexpected and surprisingly little known outcome is that Lehman Brothers International – the London based subsidiary of Lehman Brothers that handled most of their international business – actually turned out to be rather well capitalised. As revealed by the most recent report of the joint administrators (PWC, 2016) the value realised from Lehman Brothers International assets substantially exceeded its liabilities. Once all the legal and administrative costs of administration are covered and all creditors are repaid, there will still be a remaining surplus of between £6.57bn and £7.79bn pounds. The remaining uncertainty concerns ongoing legal process about the interest payments to senior creditors (the so called “Waterfall” cases). In the event that these remaining claims by senior creditors are rejected in the courts and the higher surplus of £7.79 applies, then all unsecured creditors to Lehman Brothers International will not only receive their original claims paid in full, they will also receive in full the statutory interest rate due to unsecured creditors under UK law of 8% per annum compound from the date of the administration in 2008 until final settlement, and this will still leave a further £0.4bn to be paid to equity holders (the US holding company). These high recovery values from Lehman assets explain why, as reported by (Arnold, 2014), unsecured claims on Lehman Brothers International have been trading at well above par (at £1.40 to the pound or more in early 2014).
The final payout to unsecured creditors now looks likely to be substantially higher still.

This does not mean that Lehman Brothers as a whole had a substantial surplus of assets over liabilities. (Fitzgerald, 2016) reports that by June of 2016 the court-approved payouts to unsecured creditors of the US parent company Lehman Brothers Holdings amounted to 38¢ in the dollar (with unsecured creditors in a number of Lehman Brothers Holdings US subsidiaries receiving higher payouts). Even allowing for further such payments expected in the future (the full resolution of Lehman Brothers Holdings is expected to still several more years) these unsecured creditors on the US holding company are unlikely to have their original claim repaid in full or get any compensation for the lengthy period during which these claims have been trapped in the bankruptcy process. Still an important magnifier of losses was the disorderly process of resolution under US Chapter 11, made more difficult by lack of information on assets and liabilities. Arguably (Fleming & Sarkar, 2014; Kapur, 2015), under the Chapter 14 arrangements for resolution of financial institutions introduced by the Dodd Frank Act, unsecured creditor losses would have been far smaller and it is possible that Lehman Brothers Inc. might even have been able to return something to shareholders.

The protracted process and high administrative and legal costs of resolution, the great uncertainty about both the valuation of Lehman assets and the determination of claims on those assets, firesale losses in the disorderly period that followed the initial failure and the relatively high value of remaining assets that were disposed more gradually compared to the fears of creditors and other market participants at the time of the initial failure highlights the how lack of information about assets and liabilities and the resolution of disputes over contractual obligations can contribute substantially to market uncertainty and consequent financial instability.

A second illustration comes from the problems associated with the implementation of the commitment made at the 2009 G-20 Pittsburgh summit that “OTC derivative contracts should be reported to trade repositories” (G20, 2009, pg 9). Subsequent law and regulation (e.g. the Dodd-Frank act in the US, the European Markets Infrastructure Regulation in Europe) has required such reporting but the resulting data has proved to be all but useless to regulators. A widely cited speech from a then commissioner of the Commodity and Futures Trading Commission, identifies the reasons for this:

“Specifically, the data submitted to SDRs [swaps data repositories] and, in turn, to the Commission is not usable in its current form. The problem is so bad that staff have indicated that they currently cannot find the London Whale [a reference to the massive unauthorised trades on which JP Morgan had lost several billion dollars earlier in the year] in the current data files. Why is that?

In a rush to promulgate the reporting rules, the Commission failed to specify the data format reporting parties must use when sending their swaps to SDRs. In other words, the Commission told the industry what information to report, but didn’t specify which language to use. This has
become a serious problem. As it turned out, each reporting party has its own internal nomenclature that is used to compile its swap data.

The end result is that even when market participants submit the correct data to SDRs, the language received from each reporting party is different. In addition, data is being recorded inconsistently from one dealer to another. It means that for each category of swap identified by the 70+ reporting swap dealers, those swaps will be reported in 70+ different data formats because each swap dealer has its own proprietary data format it uses in its internal systems. Now multiply that number by the number of different fields the rules require market participants to report.

To make matters worse, that’s just the swap dealers; the same thing is going to happen when the Commission has major swap participants and end-users reporting. The permutations of data language are staggering. Doesn’t that sound like a reporting nightmare?” (O’Malia, 2013)

As described in (Chan & Milne, 2013) an important step towards making trade repositories more useful for clarifying exposures in OTC derivatives markets and hence monitoring systemic risk has been the requirement that all such reporting, worldwide, make use of a common legal entity identifier the global LEI whose operation is now governed by the Swiss charitable foundation the Global Legal Entity Foundation. This however only addresses one aspect of the required challenge, ensuring that counterparties to derivative trades are uniquely and unambiguously identified, something that would have been useful for the resolution of Lehman Brothers and firms that failed during the 2007-2008 financial crisis.

Orderly resolution, carried out over a short time frame of a few days or weeks; or simply the challenge of monitoring of exposures in order to anticipate which institutions are severely (or critically) undercapitalised and so require regulatory intervention to prevent or prepare for failure, require aggregation of many thousands of different exposures. Assets, off-balance sheet exposures, debt and derivative claims should all be aggregated together relatively quickly and automatically, counterparty by counterparty, in order to get a reasonably accurate picture of net worth and the ramifications of failure. This in turn requires standardisation of data fields and data definitions across the industry. Agreed standard indentifiers for counterparties are necessary but not sufficient for addressing all the data problems that can increase financial stability risks. But, as discussed in (Houstoun, Milne, & Parboteeah, 2015; Milne & Parboteeah, 2015), financial services industries find it very difficult to agree on such data standards; while public authorities lack the detailed knowledge of business process required for them to impose such standards.

These two examples illustrate the importance of data standards and information to financial stability. The papers on this theme presented at the 2014 conference were a mixture of policy analyses and research. Two papers presented at the conference (not submitted for this special issue) addressed the trade-offs and constraints in the collection of regulatory data. (Bholat, 2016) discusses some of the challenges faced by the Bank of England, in combining regulatory reporting from a variety of different sources in carrying out its responsibility for collecting and analysing data on
prudential risk. Regulatory reporting systems are costly to set up and often determined by historical constraints that may no longer be applicable today. The still dominant conventional approach to statistical data collection is based on period collection of data in multiple forms structured like standard financial statements, designed for specific regulatory purposes and using inconsistent aggregations of the same underlying granular data. The response to the global financial crisis of 2007-2008 saw a dramatic increase in the number of such forms collected by the Bank of England, from less than 40 to more than 140, imposing substantially increased costs on the reporting entities required to complete each form, and often requiring multiple reporting of essentially the same item (for example retail deposits) using slightly different criteria for aggregation. A pilot study at the Bank has explored a different approach based on granular rather than aggregated data, focused on three of these reporting forms. This has demonstrated the possibility of considerable efficiency gains from a new approach to statistical reporting based on agreed standards for underlying granular data (a ‘common statistical language’). Having recorded or collected the underlying data, either the reporters or the Bank can then use automated aggregation to fulfil the requirements of all three different forms rapidly and at low cost.

(M. Flood, Katz, Ong, & Smith, 2013) address a different challenge, the inherent trade-off between transparency and confidentiality in regulatory reporting. While underlying data is indeed granular, not aggregated, it is also usually confidential. Firms can suffer losses from public revelation of information on their assets, liabilities or trading positions. Concerns over privacy limit public access to financial information on individuals. Furthermore, while individual regulatory agencies may be given privileged access to confidential information, they are limited by law on the extent to which this information can be shared with other public agencies. The conventional solution is for the release of the data to be restricted to aggregation across multiple entities at a level which makes it impossible to back out underlying individual contributions (note that this is an aggregate of aggregates because the data reported by each entity is already an aggregate of granular level data across individual exposures). No sharing of underlying data between agencies is possible unless expressly legally permitted. The consequence is that confidentiality limits the transparency of the financial system, for example the ability of regulators to monitor counterparty risk and the potential for a systemic contagion from the failure of a major counterparty such as Lehman Brothers is limited by constraints on the collection and sharing of individual counterparty exposures. Developments in cryptography over the past quarter of a century are though providing powerful tools that can allow the sharing of data, and flexible aggregation of underlying granular information, without breach of confidentiality hence alleviating the trade-off between transparency and confidentiality. The paper considers two such approaches. Secure multiparty computation protocols can emulate a completely trusted and trustworthy third party who can be relied upon to access confidential data from different sources and then perform computations and share final results in any desired form consistent with confidentiality. Techniques of ‘statistical privacy’ provide mechanisms for ensuring that confidential information on underlying entities is not unintentionally divulged when sharing anonymised data.

Nowadays regulators and researchers are increasingly accessing and analysing granular level data, rather than conventional statistical aggregates, often very large
data sets with multiple dimensions and thousands or even millions of underlying units. Computing power grows apace and it is possible to apply conventional econometric techniques to the resulting large panels of data. But still it is all too easy for the users of data to be led astray by the magnitude of data they are trying to digest. Three papers presented at the conference, two of which are published in this special issue, explore the use of visualisation tools for understanding large scale data sets and monitoring systemic financial risk.

(Paddrik, Haynes, Todd, Scherer, & Beling, 2016) consider the challenge of analysing data on the millions of contracts traded each day on the world’s exchanges (more than 10 million per day on the CME exchanges alone). Visualisations they consider include: (i) ‘tree-maps’ combining information on the number of orders placed on an electronic order book and the frequency of execution of those orders by individual market participant (their Figure 3) and which can be used to group market participants according to their characteristic trading strategies (high frequency traders, fundamental buyers, fundamental sellers, opportunistic buyers and market makers, see Figures 4 and 5); (ii) order book heat maps which are useful for analysing episodes of stress such as the flash crash of May, 2010; and can track changes in market participation (Figure 6) (iii) dynamic order-book animations and graphs that trace variations in market liquidity over time (see Figures 7-11). They then review some illustrative applications, including market oversight, simulation of changes in exchange rules and enforcement of regulatory requirements.

(Sarlin, 2016), in this issue, reviews the use of visual analytics and visual interfaces as a tool for monitoring systemic financial risk, emphasising in particular the requirement for effective communication of risk to decision makers responsible for implementing macroprudential policy. His discussion distinguishes four different dimensions of underlying granular data that must incorporated into the analysis: the three familiar dimensions of most economic data, by entity (such as firm or country), by time, and by instrument or economic indicator; and an additional fourth dimension required for macroprudential analysis, namely the network of counterparty linkages. He then presents a number of approaches to visualising these different aspects of systemic financial risk, for example a ‘self-organizing time map’ of the evolution of macro-financial conditions in the US (Figure 6). The real value-added though comes not from such fixed aggregated representations, but from the possibility of using web-based interactive interfaces, which allow exploration of the various dimensions of underlying data. This possibility can be explored by readers using the supporting software linked to the article.

(M. D. Flood, Lemieux, Varga, & Wong, 2016), in this issue, discuss how data visualisation can help regulatory and supervisory authorities fulfil their responsibilities, such as identifying possible sources of financial instability, monitoring systemic risks, setting and enforcing regulatory rules and promoting transparency of market information. They provide a wealth of illustrations of how visualisations have been used to explore systemic financial risk, capturing inter-alia geographic variation, network linkages, and time variation of both market pricing and potential systemic vulnerabilities (Figures 1-13). They then discuss ‘visual analytic systems’ i.e. how visualisations can be best constructed to combine the visual and cognitive abilities of human users with the rapid and automated computerised processing of data in order to provide users with effective insight into complex
systems such as the financial industry. Visualisation of this kind is an important complement to econometric and statistical analysis, providing insight into connections and developments in the data that would otherwise remain hidden and clarifying the conceptual framework that policy makers and practitioners use to understand their domains of responsibility.

While shortcomings in data obstruct the resolution of failing firms and limit the oversight of many markets, there are some rich granular level data sets available to regulators and researchers for the monitoring and study of financial stability. Two papers in this issue exploit such data sets, while a third investigates the challenge of combining multiple sources of data to assess systemic financial risk.

(Batiz-Zuk, López-Gallo, Martínez-Jaramillo, & Solórzano-Margain, 2015) use the detailed bilateral data on the interbank exposures reported to the Banco de México, in order to model contagion in the Mexican banking system following an individual bank failure. Because they have access to data on bilateral exposures they can avoid the potentially distorting impact of using ‘maximum entropy’ or other approaches used in similar studies for other jurisdictions, required to estimate unreported bilateral exposures from aggregated balance sheet information. They review the literature on interbank contagion, distinguishing older static models that consider only the mechanical ‘domino’ impact via capital losses from counterparty exposure and more recent analyses that takes account of strategic and behavioural responses and the impact on funding and liquidity. They also highlight the importance impact of network topology on contagion dynamics. They then develop their own contagion model allowing them to analyse how large exposure limits for systemically important institutions affect the stability of the Mexican banking system following an individual bank failure.

(Heath, Kelly, Manning, Markose, & Shaghaghi, 2015) use the BIS Macroeconomic Assessment Group on Derivatives (MAGD) data set on the 2012 positions in global OTC derivatives markets of the 43 largest participant banks. The issue they investigate is the impact of the introduction of a central counterparty or CCP for the multilateral netting of exposures on the stability of the market in the face of an extreme price shock. They focus on four alternative scenarios for the introduction of a CCP and the implications of this choice for network topology and the stability properties of the network, demonstrating that if the CCP itself has insufficient committed resources and therefore is itself vulnerable to failure then its introduction may increase rather reduce systemic risk (Tables 7 and 8). They also investigate a variety of stress scenarios, to establish how contagion might occur even with CCP multilateral netting in response to particular shocks. The outcome of the reported “extreme but plausible” scenarios are consistent with the network analysis in that the possibility of contagion emerges but only when there is a relatively low buffer of Tier 1 capital that can absorb losses.

(Oet, Gramlich, & Sarlin, 2016) combine a battery of different measures of US financial risk, based on 40 different time-series at monthly, weekly and daily frequencies for multiple markets from 1976 to 2013 and a further 31 time series from 2000 to 2013. They consider both coincident and early-warning measures of systemic financial conditions and assess them from the perspective on the literature on ‘binary classification’ i.e. the distinction between true identification (episodes
systemic concern indicated by policy interventions in financial markets) and false identification (no systemic concern and no policy interventions) and the trade-offs between Type 1 and Type 2 errors in identifying and predicting these episodes. They find using their full sample of data that coincident identification is improved by using multi-dimensional signal extraction using data from several markets and by employing measures that measure the severity, persistence and pervasiveness of volatility in these markets (see Figure 3 and Tables 2.-3.). They find that coincident identification is more stable when based on levels of indicators and not changes in indicators, but there is some additional information to be obtained from very rapid changes in indicators.

Their indicators are somewhat less useful when employed for early warning. Using a Box-Jenkins approach suggest that their standardised coincident measures are random walks over the period 2000-2007. Splitting their sample into two periods, a training sample from March 1990 to February 2004 and a forecasting sample from March 2004 to February 2014 they find that while forecasting ability actually improves slightly in the forecasting sample, forecasting ability is relatively weak (based on a variety of standard metrics of binary classification) and that the ranking of different indices changes markedly over the two periods serve as early warning predictors of crises.

The two remaining papers in this special issue are models of financial stability risks. (Ampudia, Van Vlokhoven, & Żochowski, 2016) employ cross-sectional data on household income and wealth from the Eurosystem Household Finance and Consumption Survey (HFCS) to develop a framework for stress testing of household balance sheets in the Euro area. Their goal is to quantify the impact of hypothetical adverse shocks on debt servicing capacity, distinguishing groups of households or countries e.g. those with the highest ratios of debt to income particularly vulnerable to shocks. A number of previous papers have used cross-sectional data to identify financially vulnerable households, using as a criteria either debt-service to income ratios or measures of ‘financial margin’ (the proportion of income left after debt service and essential household expenditures).

An advance on previous contributions of this paper is to combine information to obtain a measure of distress that used data both on solvency (as measured by financial margin) and on holdings of liquid assets. The criteria for anticipating default is that the financial margin is negative and that the household has insufficient liquid assets to cover its shortfall on financial margin for a given number of months, $M^*$. This paper also goes a step further than previous research by relating the cross-country information from the HCFS to time series of non-performing loans in the Euro area, calibrating $M^*$ on a country by country basis to achieve the best predictions of aggregate loan default in each country (Table 2 reports the resulting calibrations, which exhibit substantial variation from one country to another). Having conducted these calibrations they are then able to develop country level stress tests for the impact of three shocks – an interest rate shock, an asset price shock and an income shock – on household distress and on exposure at default (EAD) and loss given default (LGD) following the shocks (see Tables 5 and 6). To give some flavour of the outcome, they find that for most countries, but not all, a 300 basis point increase in loan interest rates has a more adverse impact than a two standard deviation loss of income (an exception is Greece which exhibits especially large
vulnerability to income shocks). While susceptibility to shocks is contained for the Euro area as a whole, some individual countries are much more exposed than others.

(Aymanns, Caccioli, Farmer, & Tan, 2016) is the one paper in this issue that does not directly exploit novel data or address issues of data analysis. They instead develop a simulation model of the dynamics that arise from variations in bank leverage driven inter alia by changes in market prices (the leverage cycle highlighted by (Geanakoplos, 2010) in which higher market prices or reduced market price volatility relax value-at-risk constraints on market participants and lead in turn to greater risk taking or willingness to invest in response to price declines). This outcome is a positive feedback with further increases in market prices or reduction in price volatility. But the leverage cycle can rapidly reverse following a shock with the same positive feedback reducing prices or increasing market price volatility. Their model provides a quantification of these dynamics in a simple model with a single risky asset and two active agents. The first of these is the bank who forecast future volatility from the recent history of market prices and then use this estimated volatility (a ‘value-at-risk’ rule) to determine a target degree of leverage towards which it gradually adjusts its portfolio. This behaviour is inherently destabilising and so the model requires also a second fundamental based investor who buys or sells the risky asset according to whether prices are below or above long run levels.

This simple set up supports a wide range of different dynamics, depending on the values of a few parameters, of which the most important are: the relative size of the bank and the fundamentalist investor; the average leverage of the bank (which depends on a long term equity target); the magnitude and direction of the relationship between anticipated market price volatility and target leverage; and the volatility and degree of clustering of exogenous shocks to market prices. Depending on the value of parameters, the underlying deterministic dynamics are (i) stable (smoothly correcting to long-run equilibrium); (ii) locally unstable with chaotic but close to cyclical behaviour within a region of attraction; or (iii) globally unstable. Introducing exogenous shocks to market prices, for the case of local instability which arises when the authors calibrate parameters to data, and solving numerically the model produces relatively lengthy cycles, with rising prices for periods of 10-15 years followed by market price crashes.

They go on to examine the impact on dynamics of choosing between two different degrees of sensitivity of target leverage to market price volatility: one the case of high risk sensitivity when target leverage is close to a negative inverse function of the standard deviation of market prices (this can be interpreted as a Basel II risk sensitive capital rule); the other a constant leverage target independent of risk (this can be interpreted as a Basel III leverage requirement). When the system is globally stable (parameters that put it clearly in region (i) so that systemic externalities are unimportant to aggregate dynamics) the volatility of the system is minimised by a risk sensitive capital rule. When the system is locally unstable (parameters that put it clearly in region (ii)) then the volatility of the system is minimised by constant leverage.

While this model, unlike other papers in this issue, does not exploit detailed granular level data, there is a close connection to the overall theme of ‘data standards,
information and financial stability’. The rich parameter dependent dynamics that arise even in this relatively simple setting are typical of the non-linear dynamic modelling widely used in the physical sciences. Such non-linear modelling requires the use of tools of analysis such as Lyapnov stability analysis, routinely applied in the physical sciences but rarely applied by economists. Such models are of considerable interest because they are able to reproduce the boom and bust dynamics of financial cycles which cannot be easily captured by conventional linearised economic models. Why then is the economics profession so reluctant to entertain non-linear models of this kind which can explain observed aggregate dynamics, instead preferring the ‘safe harbour’ of linearization even though this is inconsistent with observed aggregate behaviour? One explanation is lack of detailed granular data that can be used to calibrate and test non-linear dynamic models. Without such micro-level data non-linear specifications simply have too many degrees of freedom to make unambiguous predictions: different modelers can produce different aggregate outcomes simply by choosing different but largely uncheckable assumptions about the behaviour of individual agents.

Collectively, what do we learn from these papers about the relationship between data standards, information and financial stability? The task of promoting financial stability involves anticipating and mitigating systemic financial risk i.e. ensuring that when, as will inevitably sometimes happen, households or firms or institutions fail or default on their contractual obligations, these shocks are not amplified by further failures or by the breakdown of entire financial markets. Unlike the related macroeconomic challenge of modelling the impact of monetary and fiscal policies, these tasks cannot be carried out through focus on macroeconomic aggregates alone or modelling based on the assumption of representative agents for each sector. The collection of detailed granular microlevel data, the effective modelling of the relationships amongst this data to obtain insights into systemic financial risks, and the use of this data to support orderly failure of financial institutions are therefore essential if we are to do all we can to limit or prevent future episodes of financial instability.

The papers in this special issue and the other contributions referred to in this introduction provide an initial start on a major research programme. They discuss some of the challenges of data collection. They show some of the insights that can be drawn from analysis of detailed microlevel data and how these insights can be applied to promoting financial stability. They demonstrate the kind of modelling of financial stability that could be pursued with proper access to comprehensive granular data. But perhaps the most important lesson is the need for industry wide data standardisation. With modern information technology, the costs of data collection have fallen dramatically, allowing policy makers and practitioners access to huge quantities of data on financial firms and financial exposures. But only relatively limited progress can be made in better understanding of systemic financial risk with unstructured and inconsistent data and without the identifiers that can be used to link different data elements together.

An analogy can be drawn with the foundation of modern systems of national accounting, following the great depression of the 1930s (described for example in (Kendrick, 1970)). National accounting is the data foundation for macroeconomic modelling. Effective financial stability modelling requires its own data foundation. But
the modelling challenge today is more complicated than that faced by macroeconomists in the 1930s. In today’s world of international financial markets the required foundation is one of global financial accounting. The need to track individual exposures means moreover in turn requires this global financial accounting must rest on standardisation of financial data at the most granular level, that of individual contracts and individual loans. A great deal still remains yet to be done in order to create this required data foundation for effective financial stability analysis.

References


