## Battle for future of Data ID Standards hots up

Allan D. Grody, President, Financial InterGroup Advisors

20 January 2016

**Debate over instrument identifiers future misses the possibility of new approach**

Bloomberg, Markit, CUSIP Service Bureau and other data vendors have been posting on the various blogs and giving interviews about the administration of the International Securities Identification (ISIN) numbering system. Bloomberg and Markit offer their own codes while contending ISINs are a poor chose as a unique product identifier (UPI) for Europe's MiFid II/MiFir implementation which affects all tradeable contracts and instruments. CUSIP thinks otherwise, being the largest issuer of ISINs. ISDA would like to see the completion of regulators consultations on all identifiers before deciding.

Bloomberg describes quasi monopolistic practices carried out by embed intermediaries in the ISIN process that is designed to assign codes to identify financial instruments and contracts. They point out that newly regulated OTC derivatives are not covered by ISINs and speculate whether ESMA, the proponent of MiFid II/MIFIR legislation and its support of ISINs haven't been misinformed about the ISIN system.

While archaic, ISINs have become fundamental to the functioning of the industry's infrastructure even though new regulatory demands and teachings from ours and other industries show other, more efficient paths that can followed. Those paths, by borrowing methods from commercial barcodes and Internet domain naming procedures, can accommodate a more efficient error free automated process, one that removes intermediaries from the process and provides direct from-the-source data.

Following a half century of sovereign countries' independently constructing and assigning their own codes, and after nearly three decades of the ISIN standards, it may be time to question whether these codes, intended purely for identification purposes, are still fit for purpose. Also, whether the means by which they are assigned and registered is still valid.

This reexamination should be considered in light of more direct at-source methods of assigning and registering codes. Also to rethink the structure of these codes to enable direct data aggregation of the financial transactions in which these codes are embedded. Data aggregation has emerged as the key requirement for enterprise-wide risk management and global systemic risk analysis, and for aggregating swaps trade repository transactions and observing counterparty risk.

This rethink is necessitated not only as global regulators and government standards setters demand data aggregation capabilities but also as a vastly changed financial industry has evolved from specific coding

schemes used in sovereign markets to ones needed for global markets. It is also playing out against the background of the legacy systems of the computer age of a half-century ago, when these numbering schemes were first invented, to the real-time needs of an evolving industry's infrastructure grappling to transform itself in the digital age.

**Testing New Data Standards in the OTC derivatives Markets**

The first real test of global coding schemes is playing out in the newly regulated global swaps markets where the ISIN standard is being questioned as the appropriate standard for the required Unique Product Identifier (UPI). Also remaining is the question of whether the new financial market participant code, the legal entity identifier (LEI), can be used for data aggregation and, also, as the prefix for the unique transaction identifier (the UTI), whether in its full version (20 characters) or in its shortened (hashed) version of ten (10) characters as proposed by ISDA.

The decades old ISIN system of securities identification is administered by over one hundred National Numbering Agencies (NNAs) and Substitute National Numbering Agencies (SNNAs), and two International Central Securities Depositories (ICSDs), Clearstream and Euroclear Bank. The codes they construct define the principle method computers access, process, associate and aggregate securities transactions throughout the automated plumbing of our global financial system. That method is data mapping, itself an archaic error prone and costly process.

In addition, the Association of National Numbering Agencies (ANNA) organizes its NNA members to feed locally registered codes to a centralized data base administered by the ANNA Service Bureau. Six Financial and the CUSIP Service Bureau manage and administer the ANNA Service Bureau under contract to ANNA. In addition, the data base includes market and asset classification codes and, most recently, the Legal Entity identifiers (LEIs). These codes collectively identify issuers of securities, lead and individual fund managers, swaps dealers and swaps market participants, clearing organizations and central securities depositories (CSDs).

**ISIN vs. GLEIS**

Bloomberg has compared practices of the evolving Global Legal Entity Identifier System (GLEIS) favorably to established ISIN practices. The GLEIS is the global system established by the G20's newest standards setting body, the Financial Stability Board (FSB), to administer the Legal Entity Identifier (LEI) – a new code for financial market participants. In Bloomberg's postings the administrative process of the GLEIS is found to be more virtuous than those of the ISIN system; being more transparent; having non-exclusive assignment of facility operators; encouraging competition; and having a less restrictive (non-profit) revenue model to mitigate costs to end users.

Common to both is that the GLEIS like the ISIN system relies on local numbering conventions and independently administered facilities operators, NNAs in the case of ISIN's and LOUs (Local Operating Units) in the case of the GLEIS. In light of well-established self-assigned and self-registered methods of direct data input and code identification in other industries the necessity of intermediaries should be questioned in the value chain of code identification and assignment.

For example, both the GLEIS and the ISIN system have a myriad of companies and government agencies as data intermediaries, as either NNAs, SNNAs or LOUs, and some operate in all three capacities (WM Datenservice and the London Stock Exchange are two such examples). They include central bankers,

patent offices, data vendors, central depositories, business registrars, exchanges, clearing organizations and technology companies. These diverse set of organizations operate for-profit, cost recovery, and non-profit businesses, some operate under all three business models.

The separate NNA codes, constructed differently by each NNA, many relating to the same issuing company, are brought together through the ISIN system by adding the home country of the issuing company's two character ISO country code to the local home country code, then calculating and adding a check digit. It is then left to a mapping process to associate all the separate codes for the same issuing company together through associating them with the ISIN. In turn, as the end objective for use of these codes, all financial transactions containing these codes are aggregated into a total position of the same security.

An example of local NNA created codes, including a related ISIN is presented below for IBM's common stock and its ISIN US4592001014 created as [US (CUSIP) 4]:



**Multiple IBM Common Stock Global Identifiers**

- CUSIP (US) 459200101
- AUSTRIA 851399
- COMMON CODE 9703799
- ISIN US4592001014
- ITALY 550304
- JAPAN 584006000
- NETHERLANDS 45480
- SEDOL 2005973
- SEDOL: CANADA-TORONTO 201382
- SEDOL: FRANCE-PARIS 5217689
- SEDOL: GERMANY-FRANKFURT 5199204
- SEDOL: JAPAN-TOKYO 6003649
- SEDOL: JAPAN TOKYO 6464956
- SEDOL: MEXICO-MEXICO CITY 2667715
- SEDOL: NETHERLANDS-AMSTERDAM 4463353
- SEDOL: NETHERLANDS-AMSTERDAM 5199323
- SEDOL: PERU-LIMA 2436517
- SEDOL: SWITZERLAND – SWISS S.E. 4514325
- SEDOL: UNITED KINGDOM- LONDON 40868
- SEDOL: USA – NEW YORK 2005973
- SICOVAM 12964
- SVM 9254608
- VALOR 941800
- WPK 851399

The process of managing ISIN and local NNA codes finds analysts at National Numbering Agencies struggling with interpreting offering documents. These analysts interact with people at issuers, investment bankers, talk with lawyers and accountants, all who had a hand in creating the enabling documents (i.e. trusts, prospectuses, collective investment regulatory filings, etc.) that describe these securities. This is necessary to help them interpret the offering documents so they can describe the bond, or equity issue or mutual fund, and assign it a code. Analysts at data vendors, likewise, struggle at a data element level to interpret these documents so they can provide deeper detail beyond the name of the contract or instrument or the legal entity.

In stark contrast, most annual financial reports prepared for regulators are transformed at-source from paper documents into computer readable format through the XBRL (extensible Business Reporting Language) data tagging convention. Also, Municipal bond data is transformed from paper documents by lead underwriters through an input template at the elemental data level for setting up data processing

attributes for recording by the US's central depository, Depository Trust Company (DTC). International Swaps and Derivatives Association (ISDA) transforms contracts into data fields within an fpML (Financial Product Markup Language) data tagging protocol.

However, data from prospectuses, trusts, offering memorandum, articles of incorporation and other such documents that are the source of contract, instrument and financial market participant onboarding information is left to multiple interpretations by regulators, data vendors, NNAs and LOUs. Providing computerized formatted data, uniquely identified and in standard form, from its source would go a long way to improve data quality, eliminate unnecessary supply chain intermediaries, and reduce the overall costs of systems and manual process that supports the duplicate plumbing of financial systems in each financial institution and across all financial market utilities. We and others have estimated that cost to the industry at $50 – $100 billion.

Now, while comparisons of ISINs to LEIs are warranted, it is a very insular comparison made by practitioners historically anchored in the system of the NNA and ISIN codes, and now extended to the thinking that went into the new GLEIS with its LOU intermediaries and LEI "dumb number" codes. There are differences as noted by Bloomberg, but at the margins, with much in common. They are, in essence cut from the same 'legacy' cloth.

**Data Mapping - Compensating for Non-standard Identification**

Mapping issues of ISINs and their local codes are the same as those of the mapping issues of Legal Entity Identifiers. LEIs, like ISINS are constructed from local codes and, like ISINs, a prefix is attached. In the case of ISINs a country code is attached and with LEIs a code for the facility operator assigning the local code, known as the Local Operating Unit (LOU), is attached. An example of a LEI code is below.
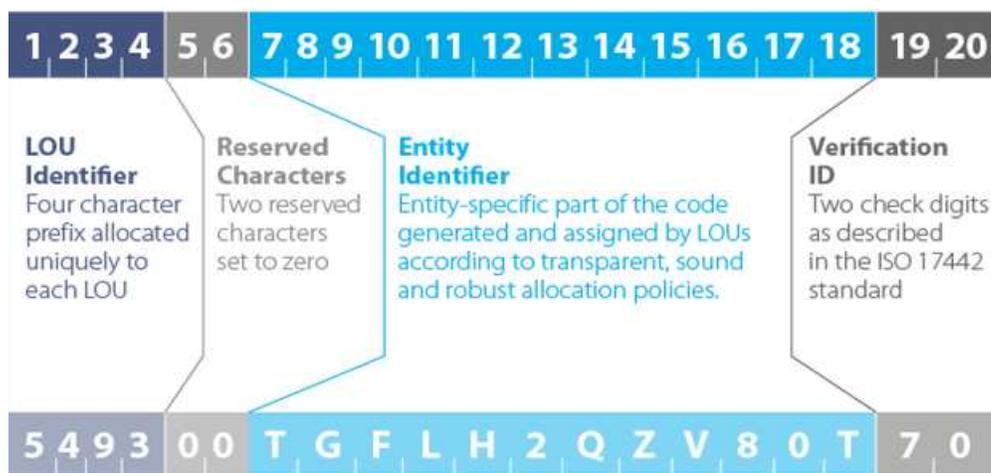


In invoking the current thinking on cryptographic discipline, not necessarily what the financial industry needs, regulators have declared the LEI codes 'dumb' numbers, although they have some structure to them. There is nothing in the LEI code that allows it to be associated with its registering parent. The LEI code is based on the ISO standard that describes 18 alphanumeric characters followed by a two digit check sum. The structure shown above was imposed later by the FSB. The GLEIS is still a work in progress. Still to be resolved is the way these LEIs are to be related to one another to allow data aggregation up through the hierarchy of ownership and control.

The ISIN code and the CUSIP number in the US, both developed long ago is the exception to the 'dumb' number discipline. The ISIN contains an observable two character country code and the CUSIP number has an issuer code prefix for the company, followed by an issue code for the category of stock (common, preferred, class of issue, etc.) or bond (rate, coupon payment frequency, etc.). This allows the codes to be used for identifying its home country in the case of the ISIN and its issuer (legal entity) and issue in the case of the CUSIP code.

In the example of IBM, its CUSIP is constructed using an assigned issuer number for IBM (issuer code 459200) followed by 10 for common stock – all common stock issues are identified as "10" for all issuers, followed by a check digit "1" calculated from the previous eight (8) digits. This allows for the aggregation and valuation of all common stock of IBM held in a position on the books of a financial company by using just the code. If this concept was used for all NNA codes and all legal entity codes, each starting out with a preassigned 'company prefix', then data aggregation using just the codes would be possible.  Why would we want to do this is explained below.

This CUSIP code construction is very much like the commercial codes found in barcodes where a company prefix is first assigned and then a product code affixed. Unlike CUSIP numbers, however, the company prefix for barcodes is uniquely assigned by a local facility operator from a global pool of numbers, then the product code is affixed by the company itself, not the facility intermediary as is the case of the NNA's CUSIP Service Bureau. This is a distinction with great meaning and implications for data aggregation that we discuss below.

Complex funds families, significant issuers of globally traded multiply listed securities, government issuers of sovereign debt, complex financial institutions, and complex financial market participants have multiple codes assigned to contracts and instruments, and legal entities. Many of these organizations will have thousands of these codes assigned.

### Data Aggregation – the End Objective

To aggregate transaction data for valuation, performance and risk analysis these independently derived dumb codes have to be mapped together, the NNA codes horizontally for a total picture of a single asset position and the LEIs vertically through its hierarchies of ownership and control for a aggregated view of a counterparty. These processes are the same across all financial market participants and their products. These mappings are in addition to the mappings to the hundreds of proprietary codes required by data vendors and software companies to utilize their business applications and data feeds.

If a company is given a globally unique registration code a 'company prefix' and uses it as the prefix for assigning its multitude of legal entities and its many instruments it issues or tradeable contracts it manufactures, then we have a mechanism to link them together for data aggregation. This mechanism, placed directly in the codes will eliminate overtime the additional costs of facilities operators, data vendors, software companies and so many other intermediaries now necessary to map both proprietary and "standard" codes at significant operational costs and risks to the industry.

### Stepping Away from a Legacy Mindset

Isn't it time we step away from the legacy mindset that is still acting to restrict our newest opportunity, the development of the LEI, from a restrictive view of an identification standard to one that facilitates the ultimate requirement of new uses for data standards – that of data aggregation?

Without a hierarchical construct to the basic code structure for contracts, instruments and legal entities the mapping issue that causes so much cost and risk will persist and the regulatory objective of systemic risk analysis will go unfulfilled. Without at-source automated input of data elements that relate legal terms from enabling documents into computer readable formats our industry will always suffer from data quality issues and, thus, be inhibited from ever fulfilling the promise of real-time straight-through-processing.