**Emily:** **Hello and welcome to this DerivSource podcast. I'm Emily Fraser Voigt, the acting editor of DerivSource.com.**

**In this DerivSource podcast Richard Robinson, Head of Strategy and Industry Relations for Symbology at Bloomberg, traces the lineage of data and looks at the gaps in the data management process, which has become significant to meet all the post financial crisis regulatory reforms.**

**Although there is no single bullet to solve all the challenges, he points to the Financial Instrument Global Identifier, or FIGI for short, as an important instrument in the identification framework that can help standardise the process and overcome some of the hurdles in a cost-effective manner.**

**Here is DerivSource reporter, Lynn Strongin Dodds.**

**Lynn:** **Hi, this is Lynn Stongin Dodds from DerivSource. We are here today having a podcast with Richard Robinson, Head of Strategy and Industry Relations for Symbology at Bloomberg. I just want to say thank you very much for taking part.**

Richard: Thank you for having me.

**Lynn:** **We're going to ask you a few questions, the first question being: in your article on data lineage, you write of its importance in the data management process which has grown in significance due to the different financial reforms. Can you please explain its role?**

Richard: Sure. Well, our business is based on exchange of information, and therefore the data that underlies that information is critical to proper understanding. As data moves through processes, it changes. Other industries, especially like manufacturing, have been managing their data much more effectively, for much longer.

In some ways, it's more clear-cut when you have a physical product. Lineage is easier to track when you have a solid thing to anchor it to, that's not subject to different interpretations or contexts. A car is a car, a windshield wiper is just that; you may have different types and styles but they are easily definable. In contrast, take something like a person. In financial services, there isn't really a single person definition, because context is more important. Someone can be a legal entity or a counterparty, a beneficiary, a trader, and all those things at the same time. We need to manage that data

properly so it represents the context that thing exists in, that is the role and / or actions it takes on that influence other elements it interacts with.

So, data management in financial services is focused on ensuring we understand what data we are using at any point in time or context, that's of quality, relates to the purpose we're using it for, and that we understand not just the upstream source it comes from, but the downstream impact that data could have on downstream systems.

**Lynn: Why is managing data quality and tracking data lineage such a challenge for professionals? What have been the main obstacles?**

Richard: Bryan Sentence of Xenomorph wrote about a great concept for describing lineage—the concept of data as water. If you run a tap somewhere that you're not familiar with, that's when you typically think about that water's lineage: where it originally came from, any contaminants that could have been introduced, how much it costs for delivery, including costs that may incur due to leaks in pipes or lack of maintenance of pipes. Data follows the same principles. Data, like water, may look the same but the source changes or any of the transformation points along the way change from that source, that can have great impact on what you're actually looking at. Something like fluoride can be introduced without your knowledge to make the data better, or a stream overflows with chemicals, which washes into your source, and now that data is bad.

Also, context comes into play, so ice and water vapour are still water, but what you can do with them is vastly different.

There aren't just one or two obstacles that you can really point to on this, it tends to be more of a mix of things, depending on the different situations.

The most common example is a large organisation that has gone through many acquisitions and mergers, thus has many different models and systems to rationalise between those mergers. You also have operational silos, functional silos, different firm types, and so on. But even a small firm that isn't burdened by legacy issues can quickly find themselves dealing with bad data if they don't have the proper procedures in place to manage lineage and quality.

Lynn: **Tabb Group issued a recent report called *Building A Framework for Innovation and Interoperability* that you cited in your article on data lineage, and it says that more than 50% of firms operate using more than one security master, and nearly 25% of asset services utilise more than ten masters. What is the impact of this?**

Richard: The existence of multiple security masters is more a symptom of the overall challenges of instrument identification than a cause, in many cases. I've written about this specific issue before in regards to what I'll call the 'myth' of data centralisation.

At a very basic level, if you store the same data in two places you need to be able to manage that data and keep it in sync with each other, especially when systems or processes act on only one of those data stores, either as inputs or as function of the output. So multiple security masters exist usually because they serve different purposes and therefore end up based on different identifiers, be it tickers, internally created codes, or other propriety identifiers like ISINs, SEDOLs or CUSIPs.

Immediately, the different databases now have different contextual views of the same thing. This means that you need to implement a framework that can manage the transformation of one context, such as a ticker, to a different context, such as a SEDOL, that exists at a more aggregate level with a different set of data tied to that identifier.

Further, we need to be able to manage that transformation both ways, to maintain how you manage discrepancies in data loss or data creation that are associated with those different identifiers. You end up with trade-offs among costs, quality, flexibility, functionality, and data fitness for the need.

There isn't a single solution to all of this, but taking a metadata manger approach and using a framework like FIGI can be a much more effective method across all these dimensions rather than forcing a single identification standard for centralising all data in some 'golden copy' database.

**Lynn: To that point, how can an instrument identification framework, and in particular FIGI, support the data management process? What are the benefits, if you can maybe provide a bit more detail on FIGI.**

Richard: Sure. So I began to touch on that in the last question a little bit. FIGI as a framework based around metadata approach provides a flexible foundation for data management professionals to build a governance approach to identifying financial instruments in the right context, based on different functions or operational requirements. The basis of any proper data management approach begins with the ability to uniquely identify something with a clear, unique definition, with permanence.

So in the data world, we commonly refer to a primary key or more modern, uniform resource identifier approach. The unique key always points to the same object, so you can always find it, wherever you go.

The analogy we go to on this is part numbers for a car, or maybe a part for an appliance you need to fix. With that part number you can be sure it will fit for your particular need, because it matches, regardless of the manufacturer, or the part maker.

FIGI is the first financial data standard that has these exact qualities—uniqueness and permanence, coupled with an open data approach. Your part number wouldn't be as helpful if you had to pay for someone to first find out what the part number is, and then also pay for using that part number to order the part. So being based on the metadata approach is what makes FIGI a framework, and it makes it extensible as well as enables the core structure of a contextual and self-referencing system of identification.

You can use FIGI in many ways, from a core data management strategy across instrument masters, to middleware and metadata layer for interoperability between legacy systems and new builds that are based on different identification approaches.

**Lynn: Although that sounds very straightforward there's still a lingering misunderstanding of FIGI, and I was wondering if maybe you could explain why that is?**

Richard: The biggest challenge is always preconceived notions, and especially understanding FIGI as a framework, as opposed to just an identifier like say an ISIN. But proof is in seeing and doing. You can go to openfigi.com and access the framework and metadata without restrictions, regardless if you are a large firm or individual, or if you're a Bloomberg client or not, or even if you're a competitor.

Instrument identification space… there's nowhere else that you can do that. To even get a look into exportable or API-based data, you usually need to register and be approved for any other identifier. These proprietary identifiers like ISIN and CUSIPs have tiered services, where maybe you get a few manual searches for free, but anything useful you need to pay for, and the more you pay the better the access mechanisms you get.

In many cases, key data is in text fields versus discreet fields. What's important to note though is that existing proprietary identifiers all have a role to play. They exist because at some point there was a problem that needed to be solved, and the identifier addressed that need. So there are many legacy systems based on them, and they're not going away any time soon.

**Lynn: I was wondering if you could separately comment on—touching on your point of the ISINs—there is a debate going on and participants are split as to whether ISINs for OTC derivatives would be fit for**

**purpose, or whether there will be other types of identifiers under MiFID II. How has the debate progressed so far?**

Richard: Currently I think it's fairly clear the industry has decided that the ISIN for OTC will only be fit for the MiFID II regulatory requirements for reporting. It's being used only because ESMA has mandated it, and that was against the recommendations of the industry. Also, note that the cash markets are not serviced by the derivatives service bureau, but by the ANNA service bureau, and there are gaps in the ISIN coverage there for MiFID reportable securities as well.

But the debate is really muddled by conflicting perspectives, some vested interests, some misinformation or misunderstandings, and a mixing of requirements definition with data approach.

So if you take that last part first (the data approach), you see a similar thing sometimes with technology over the past 20 years as it became more accessible for someone to code their own webpage or GUI. Many times the result is something functional that does the job it needs to do. But without a programmer's discipline, actually doing that job instead of the user doing it, it's many times difficult to change and update, hard to integrate with other solutions, and prone to glitches, errors, and other problems with quality.

So right now, you have the involvement of a lot of highly intelligent, experienced, senior people in derivatives as users. But not so much the modern data experts on the other side, helping separate the data requirements from the underlying solution. In some cases the regulation has leapt ahead and inadvertently defined a rigid data solution as opposed to the actual regulatory need.

As a result, you have a very potentially single-purpose solution in what the MiFID II or RTS 23 defines. The danger is when users start to want to expand that usage, either through straight-up misuse of what you get at the end that's OTC derivative ISIN, or by effectively jerry-rigging a solution that's less than robust. Both paths are prone to data quality issues and have a good potential to result in errors, and unnecessary costs to maintain.

FIGI, in contrast, will be able to satisfy the industry's operational and trade lifecycle needs in this space while maintaining that back to the ISIN world for your RTS 23 reporting needs.

**Lynn: Thank you very much for your time and your insights. They're much appreciated.**

Richard: Thank you very much for having me.

**Emily:** **Thanks Lynn, and thank you Rich, for joining us in this podcast. If you would like to see the transcript of this podcast, please go to the show notes page on DerivSource.com. You'll also find related articles on this topic at the website.**

**To hear other podcasts, please go to our Podcast page on DerivSource.com, or you can download the free DerivSource app and listen on the go on your commute, at the gym, or wherever is convenient to listen to our industry interviews.**

**Thank you for listening, join us next time.**