Collibra

Article

Semantic Glossary Supports MDM Governance at Janssen Research & Development

The process of data governance most often involves the organization of policies, processes and controls across owners and users of customer, product or other data domains, often with an operational focus. Research and development data governance projects, as in the pharmaceutical industry, are no less urgent given the stakes of managing costly drug pipelines, but involve more parties and processes that are drawn out over years.

Janssen Research & Development, a captive business of Johnson & Johnson, presents one such example. (JRD's story was featured in depth at the 7th annual MDM & Data Governance Summit October 14-16, 2012 in New York.) The company develops treatments for ADHD, mental health and pain management and requires managed information with additional requirements for agility across functions and compliance.

As ever in pharma R&D, the watch words are time to market, and in JRD's setting of distributed research and scientific pursuits where opinions don't neatly fit spreadsheets, governance has benefited from the creation of a semantic business glossary. The glossary is being implemented through a collaborative governance tool, in this case from Collibra, which lets business and technical people implement and to a degree automate definitions and rules into dictionaries and workflows. "In an R&D setting there are many systems and opinions that need to come to agreement across drug development departments," says Patrick Genyn, who heads the drug development team for information governance at JRD.

For that reason, MDM governance is initiated in decentralized communities of practice that come from business, science and those charged with setting quality requirements for specific data points. The community leaders meet informally each week and build a version for their data domain in Collibra, which is then circulated in their community for feedback.

One of the biggest challenges was trying to convey what ownership meant, says Erica Gooch, the lead for drug development and information governance at JRD. "Some people were shy about being accountable for a definition or the process the data point should go through or what the quality rules for data cleansing would be." For people usually off working alone, Gooch says, it was necessary to explain that their input would not necessarily be the end all, but an informed baseline to start from that could be mapped, and, when ready, integrated across systems or organizations.

The community reps have been able to create "pretty solid versions" of rules and definitions with feedback and takeaways from their functional areas, says Gooch, which are submitted for approval to the glossary for the organization at large. That is facilitated through Collibra, which Gooch says exports and shares organized views with swim lanes for processes and identified owners, much more useful than a spreadsheet or shared collaboration repository.

"When we sit down to compare areas and processes we see what you have in yours and what we have in ours," says Gooch. "We can leverage each other's success and ultimately we'll probably update some nice points into our definition and our semantic glossary." It's easy, she says, to point the volume of work done and sources involved, which is also a good way to advertise data governance to the organization.

Genyn thinks JRD will also benefit from quicker processes and much more secondary usage of data. "At Janssen we are known for getting submissions out the door fast when somebody comes up with an idea for a new molecule. Now we can think about reusing that information in other projects with different drug compounds and indications."

Secondary usage is still a goal at JRD where the company has so far put master data into production for specific compounds and will soon have another domain for clinical research investigative sites, which are all the addresses and facilities where clinical trials take place. It's a lot of manual effort to put together, Genyn says, but the research and other information has a lot of shelf life for reuse in new developments. Once completed it will mesh easily with new data entering the system.

"The tricky part I think we have had some success with where others struggle is managing the semantics and the owners, and including our narrower and wider views from science and the business," Gooch says.

In the end, for a big organization like J&J it will be helpful to have a central place online with managed communities, synonyms, reference data and FDA regulations that can be handed along to outsource partners, like TCS, which handles profiling and validation for JRD.

It's a governance organization that makes heavy use of the semantic platform and work up front that will free data owners from repeated tasks down the road, Genyn says. "Normally the data owner gets all that ongoing work and those people are extremely busy bringing valuable research and opinions to the table. Eventually we'll have sensible views where a data owner can approve changes with a quick turnaround that's not a big burden to them."