



IT-DIRECTOR.COM

Content: a data integration niche

Published: 1st April 2008

It is typical of maturing markets that as the leading players consolidate or get acquired by bigger fish, then new companies spot opportunities in niche sectors of the market. This is exactly what has been and is happening in the data integration market with less well known companies emerging to specialise in such things as web ETL, content ETL and data migration. Note that when taken in the context of these sorts of areas, “niche” does not equate with “small”. In this article I want to focus on EntropySoft, a French company specialising in content ETL. That is, it supports extract, transform and load for documents rather than structured data.



I have to admit that I hadn't really considered content in ETL terms previously, except in the context of it being used alongside conventional data integration processes. However, if you think about it, it should be clear that there are the same consolidation, migration and integration issues across content repositories as there are across databases. However, most of the tools in this space are only point solutions. For example, if you are Documentum and you want people to migrate from FileNet to your environment then you might offer a tool that read FileNet and wrote to Documentum but you wouldn't offer a general-purpose tool that aimed to read and write from all leading document sources.

So, to be generic, you need a strong set of connectors to the likes of Documentum, FileNet, OpenText (and Hummingbird), IBM, Alfresco and so on, as well as support for Lotus Notes, Microsoft Office and SharePoint, and so forth. Moreover, you need to be able to both read and write to all of these. According to EntropySoft it offers some 25 bi-directional connectors of this type at present, which makes it, by some margin, the leading player in this market. These connectors can be (and are) licensed by OEM partners who want to build these capabilities into their own products, accessing them via a Java API.

Once you have all of these connectors then you can offer federated access to these content repositories and this is precisely what EntropySoft does, through a product called EntropySoft WebTop, which can be licensed as a stand-alone application or in conjunction with EntropySoft Content ETL.

As far as Content ETL is concerned you define jobs, schedules and process flows through Content ETL Studio, which looks and acts pretty much like the same sort of development environment you would have in a conventional ETL tool, with drag-and-drop onto a palette.

However, it isn't the details of how EntropySoft works that are important: it is the fact that it does so at all. I don't know any of the leading data integration vendors that have paid much attention to content. Yes, they have support for unstructured data but they don't have connectors that have any knowledge of the (metadata)

formatting that content management vendors impose on content, which is precisely what EntropySoft does have. If you have any sort of migration, integration or consolidation project involving substantial amounts of content then you should give serious thought to deploying EntropySoft as a part of your solution.

Philip Howard
Director of Research, Bloor Research



© Bloor Research 2008