

Turning the tables

Miramo® is playing a critical role in ensuring quality and consistency in the production of PDF and print versions of complex tables that convey a wealth of economic statistics sold around the world by the OECD.

With 30 member countries representing the vast majority of the world's democratic market economies, the OECD is a leading source of global economic data, producing a wide range of publications at a variety of frequencies and in an expanding range of formats. The published documents are made available online through OECD's iLibrary and the OECD Online Bookshops.

"We have more than 15,000 customers buying OECD publications, and a readership in the hundreds of thousands, since many of the customers are institutional librarians buying on behalf of their community of users – ministries, universities, businesses and so on," explains Toby Green, head of the OECD's publishing division.

A large proportion of the documents generated and sold by OECD contain data on everything from the price of mobile phone

"Miramo allows us to generate complex statistical tables automatically without touching the XML data, guaranteeing the quality"

calls to migration statistics. Presently there are some 50 statistical publications, varying from 150 pages in frequent publications such as *Monthly Statistics of International Trade*, to 500 pages or more for annually updated documents like the *International Trade by Commodity Statistics*. Publication frequency ranges from regular monthly and quarterly schedules to one-offs, depending on the nature of the content, but the one thing they all have in common is that they contain a lot of tabular data.

Multiple workflows

Prior to the introduction of a centralised database, the production of OECD's statistical publications was achieved via a variety

of workflows, typically involving manual assembly of data in Excel and export to PostScript for output and printing. Although this process was quite efficient, the finished results lacked visual coherence.

"There was no common look and feel," comments Marion Desmartin, Head of Production, Publishing, "and we had no consistency across publications."

The implementation in 2005 of a single centralised database, called DotStat, for OECD's statistical information allowed an XML data stream to be output. At this stage, Miramo was chosen for its sophisticated and highly productive table formatting capabilities:

"Some of our books measure 32 by 24 centimeters, have 300 pages and are published in five versions," says Desmartin. "To prepare the tables in FrameMaker alone would have taken days, but Miramo has the capacity to process all this information and create the table layouts quickly."

However, attempting to feed this into the various existing publishing production workflows was problematic:

"It took hours of programming time to 'fix' the multiple workflows to work with the XML input and to adapt the programs to each specific publication's layout requirements. It was clear that we needed a single system and exchange format to link the data repository to the publishing unit," Desmartin explains.

"By developing a generic architecture with modules that could be adapted and applied to all the publications, any enhancement to it would benefit all of them without having to rewrite code for each."

Timely tables

As market pressure built for ever more instant communication, it was clear that the amounts of manpower and time required to produce publications quickly were an obstacle that needed to be addressed.

PROFILE

Organisation for Economic Cooperation and Development

Established in 1961, the Organisation for Economic Cooperation and Development (OECD) provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and co-ordinate domestic and international policies.

Production of a wide range of statistical publications via multiple manual workflows was leading to inconsistencies in presentation and potential quality control issues until Miramo was implemented as a key part of a bespoke Table Output Generator. This automates the production of PDF and printed versions of large numbers of complex bi-lingual tables, ensuring both consistent formatting and accuracy of content.

“Even though the OECD deals in long-term analysis rather than short-term predictions, it has become important to be able to disseminate information and data in a timely manner,” says Green.

With the help of various sub-contractors including Pierre Attar of Paris-based Tireme SA for the overall design and architecture and Bruno Chatel of Chadocs in Marseille for development, as well as VAR and Miramo distributor Antéa, OECD developed a Table Output Generator (TOG) application in 2007 to speed the production of the tabular material that forms the core of its statistical publications. As OECD was already using Miramo for its automation capabilities, it was logical to keep it as a core composition engine to automate the PDF and paper output of tables from the XML data stream generated by the DotStat data repository.

One source, many outputs

A major goal for the redesigned publishing architecture was to be able to produce multiple outputs from one source. As well as conventional printed publications, there was a growing requirement to support a variety of electronic formats for different applications. Currently, Miramo generates the tables in FrameMaker's native MIF format, from which printed versions and PDF for online use are made. Printed output is produced at an in-house print facility or via external print providers and is likely to evolve via print-on-demand solutions.

Miramo is playing a key role in achieving the project goals via its part in the Table Output Generator and as a result, a large part of the automation target has been achieved. Using FrameMaker, Miramo and the TOG, the creation of PDF pages containing the tables is now achieved in a matter of minutes.

“Creating the tables without human intervention is a huge advantage over the previous process,” comments Desmartin. “Miramo allows us to generate tables automatically without touching the data, guaranteeing the quality, something we couldn't do otherwise.”

Excel and HTML outputs are also produced, via other Table Formatter modules added to the TOG in 2008; these and PDF are usually produced simultaneously. Other output formats under consideration are EPUB for e-books and SDMX, an international standard for the exchange of statistical data.

OECD is highly satisfied with the support it receives from Antéa and Datazone, developer of Miramo.

“Datazone provides very helpful, timely and friendly responses to our enquiries. They are very flexible in providing us with special fixes, additional information, temporary licences, help with migration, and so on,” comments Technical Project Manager Jakob Fix, adding, “As the French reseller of Miramo, Antéa is our first port of call for everything Miramo.”

The image shows a screenshot of a statistical table from the OECD website. The title is '4. ... Animal and vegetable oils, fats and waxes / Huiles, graisses et cires d'origine animale ou végétale'. The table contains multiple columns of data, likely representing different countries or regions, and rows of statistical values. The data is presented in a structured, tabular format typical of OECD publications.

Future directions

Although the statistical tables represent the majority of the content of OECD's publications, this is complemented by front matter, methodological information, sources and other matter. This is usually supplied as Word files or extracted from another database that contains metadata; this content is not yet integrated into the automated publishing workflow, though there are plans to incorporate it. Miramo will be able to handle the PDF/print output options for these integrated documents.

“With tools like the TOG and Miramo we can think about even more frequent updates to our publications, too” say Desmartin.

Other benefits have come from the centralised publishing system strategy as well:

“With the initial costs amortized, production costs are reduced and staff can be deployed for other more demanding and valuable tasks,” she adds.

OECD is hoping to increase the number of its statistical publications from the current 50 to around 100 by the end of 2009; there are no technical obstacles to doing so as the automated system has plenty of capacity.

OECD currently publishes in English, French and bilingual versions, but is actively looking into supporting further languages:

“As an international organization with member countries speaking languages as diverse as French, English, German, Hungarian, Finnish, Japanese and possibly Russian and Chinese soon, we are of course interested in the possibility of providing this information in as many languages as possible,” comments Green.

Miramo is capable of supporting all these languages, including those with non-Roman character sets and writing directions, so again, the technology is already there.

Other possibilities for the expansion of Miramo's role within OECD could include the production of analytical publications for PDF and print in addition to the statistical ones. Further integration of Miramo could be pivotal in automatically transforming data from XML databases into PDF. ■

OECD publications such as the *International Trade by Commodity* report can run to nearly 600 pages and contain more than 500 bi-lingual tables. This used to have to be created in sub-sections and manually assembled to make the completed PDF, taking up to a month; Miramo produces it in one go, in less than three hours.

Miramo®

Antéa
Tel: +33 (0) 9 72 54 91 10
www.antea.fr

Datazone Ltd
Tel: +353 64 66 28964
Fax: +353 64 66 28965
www.miramo.com

© Datazone 2016