

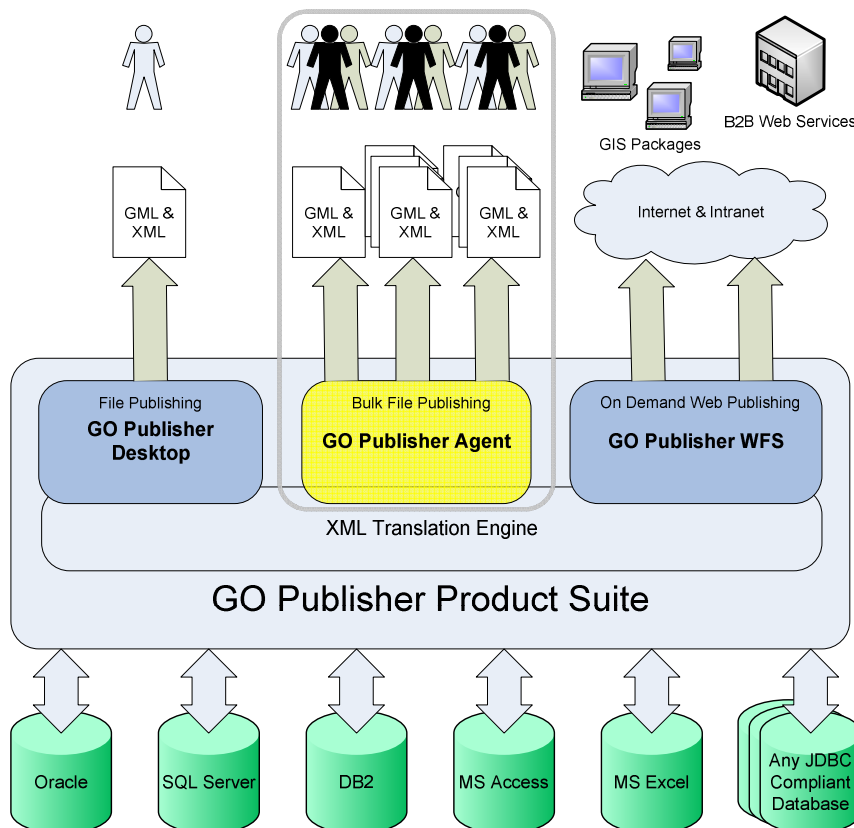
GO Publisher Agent

- for high volume, on-time GML delivery -

Publishing Information as XML/GML

XML is the most widely used format for encoding and publishing digital information, and GML is rapidly becoming the format of choice for geospatial information. XML and GML are now firmly established as internationally recognised standards, increasing the demand on organisations to publish data in these widely adopted formats. For organisations with small data volumes and a single supply pattern, this is a relatively simple task. However, for organisations with large data volumes and diverse supply patterns, the procedure can be cumbersome, complex and costly often requiring a bespoke system to be developed.

Snowflake Software's GO Publisher product suite addresses the publication of information as XML/GML from any standard database at a number of levels. At each level, Snowflake's unique schema translation technology is used to deliver information in schema that are user-defined and application-oriented and not determined solely by the requirements of the storage database.



The GO Publisher product suite supports delivery via three defined options: Single File Publishing (GO Publisher Desktop), Bulk File Publishing (GO Publisher Agent) and on-demand web publishing (GO Publisher WFS).

The file based publishing features of GO Publisher Desktop make it the ideal, entry level solution for users looking to send single file publications to one or two single customers. It provides easy-to-use 'drag and drop' functionality with a familiar graphical user interface.

Processing multiple orders to a variety of clients, however, introduces overheads that cannot be addressed by a desktop client solution. Typically, multiple users will require access to the same database infrastructure, each demanding timely reaction and delivery to their demands. Realistically, the only way to implement this is via an automated system, supplemented by override provisions for operator intervention when urgent requests need to be prioritised.

Furthermore, where the data file volume exceeds the limitations of the network or system configuration, and specifically when this contains geographical data, "chunking"¹ into manageable units may be needed to ensure timely, complete and accurate delivery of data even with single file delivery.

These scenarios are addressed by GO Publisher Agent.

GO Publisher Agent

GO Publisher Agent is a 'Standards Commercial Off-The-Shelf' (SCOTS) publishing system for bulk and background publication of XML and GML. It provides a comprehensive set of delivery options including batching, geographical chunking, file size chunking and compression. The geographical extent of a publication can be cut to a user definable order polygon.

GO Publisher Agent provides command line interfaces to the GO Publisher data translation engine, so it can be integrated into existing workflow and order processing flow-lines. Agent also provides a comprehensive queue management system enabling multiple processors to work in parallel on a single publishing task. The ability to process a single order in parallel is essential to make best use of hardware to meet the high demand of bulk data publishing. Additional processors can be added at any stage to further improve throughput times or increase capacity. This flexibility provides future-proofing against rises in demand that are impossible to predict with accuracy.

The Task Queue

The operation of GO Publisher Agent is based around the task queue. This queue contains a set of tasks that need to be carried out by GO Publisher to fulfil orders for data.

The Queue Manager is responsible for organising the content of the queue. On receiving an order, the Queue Manager breaks the processing of the order into one or more tasks and places the tasks on the queue.

The Agents are at the heart of GO Publisher and are responsible for carrying out the main and most complicated work. The Agents run continuously, taking tasks from the queue, processing them, translating data from the database and creating output XML files.

¹ The concept of Geographical and File Size chunking are covered in a separate White Paper. Please visit the White Paper section on www.snowflakesoftware.co.uk or contact Snowflake Software directly for further details.

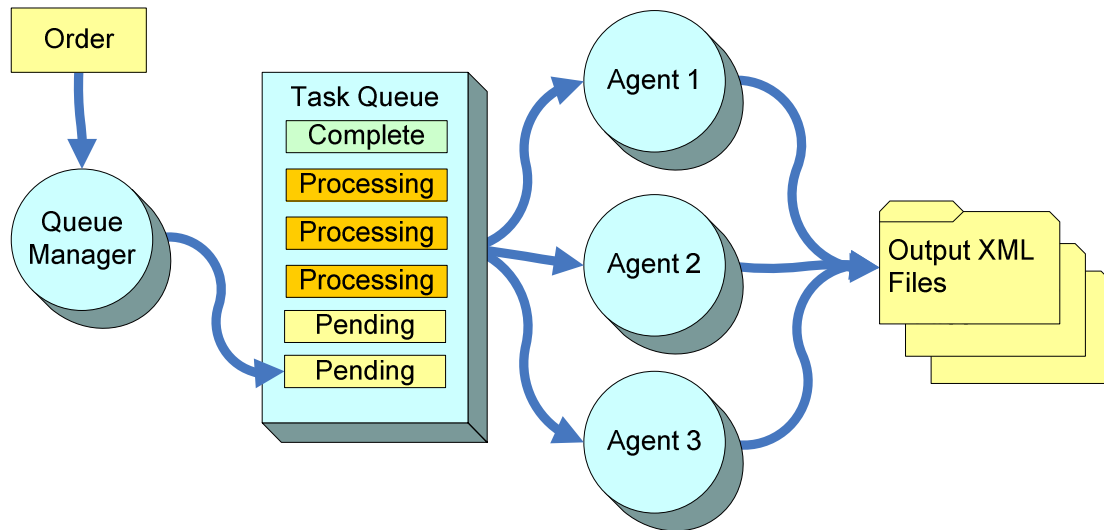


Figure 1 - The queue manager adds new orders to the task queue to be processed by the data processors. The data processors take tasks from the queue in priority order.

This Queue Manager – Agent relationship provides a number of benefits:

- It means that GO Publisher makes the maximum use of machine resources since all the Agents produce output files relentlessly until the queue is empty.
- The queue can be configured to control the order in which tasks are carried out. This ensures that orders are processed in a timely manner with urgent orders taking precedence over less urgent ones.
- It allows the client to place orders with GO Publisher whenever the requirement arises; the client never has to check whether GO Publisher is free to process the order.

Resource Management

Prioritising Orders

A priority is assigned to each order when the order is added to the queue for processing. Since the Agents take work from the queue in priority order, urgent orders can be processed as soon as possible according to their higher priority ranking. The lowest priority orders can then be run as a background task, being processed only when there are no urgent orders in the queue. GO Publisher automatically handles this rescheduling of processes without the need for user intervention or the need to cancel current tasks.

Ensuring Acceptable Throughput - Assigning Orders to Specific Agents

For normal operations, the prioritisation of orders is sufficient to ensure that all customers receive their data in a timely manner. However, if a particularly large order is received, it could cause congestion in the processing system.

Specifically: if an order has been given a high priority, it could take up all the available processing capacity and so severely delay other orders from being processed; if the order has been given a low priority, it may take too long to process if a series of higher priority orders are added to the queue.

This can be solved by assigning any specific order to a particular Agent. In this scenario, the nominated Agent will work on the assigned order in preference to any other orders and the remaining Agents will ignore this specific order. As an example, if there are three Agents, Agent 1 can be dedicated to Order 1 whilst the other two Agents carry on processing other Orders 2, 3, 4 etc as normal. This way the Order 1 is sure to be processed, but it will only be allowed to use Agent 1, so cannot cause congestion. The delivery time for Order 1 is predictable since it will be processed continuously until complete. Meanwhile, the delivery times for Orders 2, 3, 4 etc are not unduly affected.

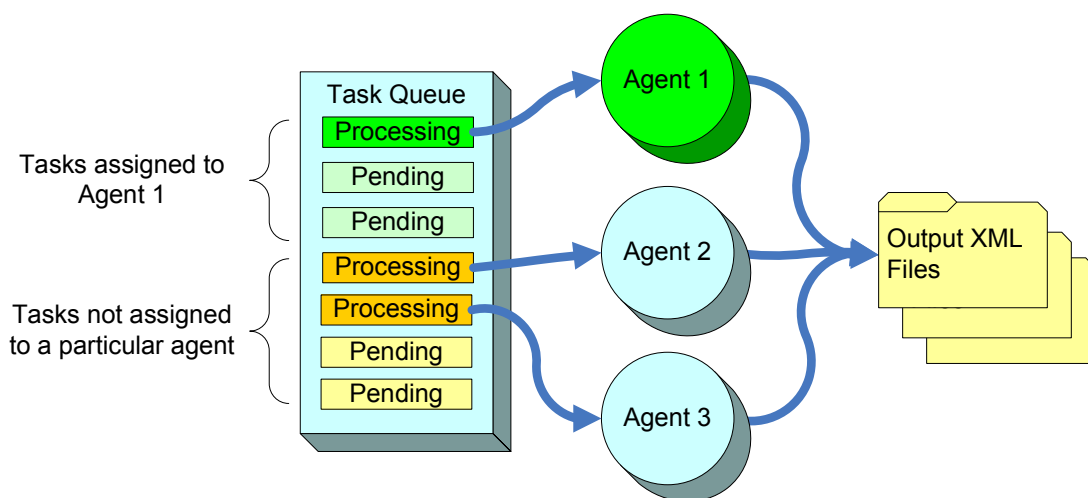


Figure 2: Agent 1 will always choose its assigned tasks in preference to others. Agents 2 and 3 will ignore tasks assigned to Agent 1 thus ensuring that the rest of the queue continues to be processed in parallel to Agent 1's assigned tasks.

Dynamic Queue Management

As orders move through the queue, GO Publisher fully manages and monitors their status. The operator can see which orders have been processed, which orders are in progress, and which orders are yet to be started. Typically, GO Publisher is programmed to run independently, leaving the system operator free for alternative projects. However, GO Publisher also supports the capability for user intervention. As such, the operator may change the priority of any order whilst it is in the queue or assign the order to a particular Agent after it has been added to the queue. This allows the operator to dynamically manage the processing of orders and react to problems with particular orders taking too long to process.

Chunking

GO Publisher Agent is designed to handle large volumes of data. Typically one order may contain too much data to be conveniently held in a single output file. GO Publisher deals with this by dividing the output between a number of files (or 'chunks'). This process is known as chunking² and can be assigned according to geographical location of a feature or by file size.

The GO Publisher Product Suite

At its simplest level, GO Publisher takes data from any standard database, translates it into the ubiquitous open standard XML and enables the user to either publish that data according to any given schema or create a bespoke XML schema to publish it to.

The core of GO Publisher is the XML translation engine, enabling on-the-fly translation of the data. Since the engine sits at the heart of the product suite, it services each of the product delivery mechanisms: the translation to XML need happen only once for single file, batch file or on-demand web publishing options. Users can start today on the desktop and upgrade to Agent and WFS as the business determines.

Features and Benefits of GO Publisher Agent

In addition to the basic flexibility and power of GO Publisher, GO Publisher Agent provides comprehensive facilities for handling high volumes of diverse data orders:

- Queuing of orders for processing to ensure maximum throughput
- Prioritised processing of orders to ensure timely delivery of data
- Allocation of individual orders to specific processors (agents) for best use of machine resources
- Dynamic management of processing to allow the operator control to react to changing priorities for orders
- Chunking of data to make large orders manageable
 - Geographical chunking for spatial data to provide chunks which can be used independently of each other
 - File size chunking for optimum transfer of whole orders
- Compressed output in both GZ and ZIP format to reduce file size output and disk requirements
- Extensibility to meet peak demands or future increases by adding extra processors.
- Ready integration into external order-processing and workflow systems

GO Publisher is designed to meet the most demanding requirements of heavy duty information publication and standards compliance. For further details on proven schemas and performance metrics, please contact "info@snowflakesoftware.co.uk".

² See separate white paper from Snowflake Software