



**British
Geological Survey**

NATURAL ENVIRONMENT RESEARCH COUNCIL

Application specifications for generating output from the DGSM

DGSM Programme

Internal Report IR/03/089

BRITISH GEOLOGICAL SURVEY

INTERNAL REPORT IR/03/089

Application specifications for generating output from the DGSM

J L Laxton

Key words

DGSM, GSF, metadata,
applications.

Bibliographical reference

LAXTON, J. L. 2003. Application specifications for generating output from the DGSM. *British Geological Survey Internal Report*, IR/03/089. 27pp.

BRITISH GEOLOGICAL SURVEY

The full range of Survey publications is available from the BGS Sales Desks at Nottingham and Edinburgh; see contact details below or shop online at www.thebgs.co.uk

The London Information Office maintains a reference collection of BGS publications including maps for consultation.

The Survey publishes an annual catalogue of its maps and other publications; this catalogue is available from any of the BGS Sales Desks.

The British Geological Survey carries out the geological survey of Great Britain and Northern Ireland (the latter as an agency service for the government of Northern Ireland), and of the surrounding continental shelf, as well as its basic research projects. It also undertakes programmes of British technical aid in geology in developing countries as arranged by the Department for International Development and other agencies.

The British Geological Survey is a component body of the Natural Environment Research Council.

Keyworth, Nottingham NG12 5GG

☎ 0115-936 3241 Fax 0115-936 3488
e-mail: sales@bgs.ac.uk
www.bgs.ac.uk
Shop online at: www.thebgs.co.uk

Murchison House, West Mains Road, Edinburgh EH9 3LA

☎ 0131-667 1000 Fax 0131-668 2683
e-mail: scotsales@bgs.ac.uk

London Information Office at the Natural History Museum (Earth Galleries), Exhibition Road, South Kensington, London SW7 2DE

☎ 020-7589 4090 Fax 020-7584 8270
☎ 020-7942 5344/45 email: bgs london@bgs.ac.uk

Forde House, Park Five Business Centre, Harrier Way, Sowton, Exeter, Devon EX2 7HU

☎ 01392-445271 Fax 01392-445371

Geological Survey of Northern Ireland, 20 College Gardens, Belfast BT9 6BS

☎ 028-9066 6595 Fax 028-9066 2835

Maclea Building, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB

☎ 01491-838800 Fax 01491-692345

Parent Body

Natural Environment Research Council, Polaris House, North Star Avenue, Swindon, Wiltshire SN2 1EU

☎ 01793-411500 Fax 01793-411501
www.nerc.ac.uk

Foreword

This report was produced as part of the data architecture and metadata sub-projects of the DGSM Framework project.

The objective of the DGSM Framework project is to define corporate strategies, structures and procedures within which geospatial models can be developed, archived and linked to associated data and knowledge.

Eight sub-projects have been defined that together are planned to fulfil this objective:

- Data Architecture
- Metadata
- DataWarehouse (Object store)
- Application development
- Software standards
- Best Practice
- Training
- Project management

These projects will be tested by studies carried out under the parallel DGSM-UK (the population of the DGSM).

Contents

Foreword	i
Contents	ii
Summary	ii
1 Introduction	1
2 GSF Trial Output Application Specification	1
2.1 Introduction.....	1
2.2 Selection module.....	2
2.3 Output and display components.....	4
3 DGSM Metadata Application Specification for Selection, Retrieval and Evaluation	6
3.1 Introduction.....	6
3.2 Data discovery component.....	6
3.3 Model discovery component.....	12
3.4 Inference metadata component	18
References.....	19

FIGURE

Figure 1: Overall application schema.....	20
Figure 2: Selection module schema.....	21
Figure 3: Overview of DGSM metadata application.....	22
Figure 4: Data discovery component application schema.....	23
Figure 5: Data retrieval module schema.....	24
Figure 6: Model discovery component application schema.....	25
Figure 7: Model selection module schema.....	26
Figure 8: Model display and retrieval module schema.....	27

Summary

As part of the development of the DGSM Framework two application specifications were drawn up. The first of these was produced by the Data Architecture sub-project and was a specification for an application to allow users to select data from the GSF database, based on spatial, metadata and attribute criteria. The selected data could then be used to generate various user-selected model visualisations and attribute displays on the user's desk-top. It was a key requirement that it should be possible to run the application over the internet using a standard browser with no plug-ins or add-ons that are not freely downloadable. The application was designed to provide general purpose access to the GSF database to a wide range of potential users outside of BGS. It was not aimed at any particular group of end-users and was seen as the DGSM analogy of the

traditional geological map. This application was designed to test all the main elements necessary in building GSF output applications as well as providing a useful deliverable in its own right. In the future the methods developed will be used to build GSF output applications to address the requirements of more specific groups of users.

The second application specification was produced by the Metadata sub-project and was designed to show how the DGSM metadata could be used. The DGSM metadata is divided into three distinct types, each designed to answer a specific type of question posed by a particular group of users:

1. Data Discovery metadata is designed to tell modellers what suitable data exist for creating a particular model.
2. Model Discovery metadata is designed to tell model users what suitable models exist for a particular application.
3. Inference metadata is designed to tell model developers and advanced users how a model was developed.

The application specification was designed to define an application, for internal BGS use, that would use the metadata to answer the questions they were intended to address.

Although the two applications are aimed at different users and have different functionalities, there are nevertheless some similarities between them. In particular the selection modules and the GSF model display component are similar. This suggests that in the future, as the difference between what it is technically possible to provide to internal and external users reduces, it may be desirable to merge the functionality described in the two application specifications in this report. For this reason it was thought sensible to bring the two specifications together in this internal report.

1 Introduction

This report describes two DGSM application specifications that have been developed. The first of these is for a trial application to enable a user external to BGS to be able to retrieve data from the GSF and display it using a standard internet connection. The second application is for internal BGS users and is designed to enable selection, retrieval and evaluation of models held in both the GSF and GLOS using the information held in the DGSM metadata.

Although developed separately to address different problems these two application specifications have areas of overlap, particularly in the selection and evaluation of models prior to display and in the display of GSF models. At present technology restrictions mean that significantly greater functionality can be provided to internal BGS users than external users, but this difference is likely to reduce in the future. It would then be sensible to combine and enhance the two application specifications presented here to create a single integrated specification covering both internal and external users.

Development of the applications is the responsibility of the DGSM Development of Applications sub-project. The GSF trial output application is currently under construction.

2 GSF Trial Output Application Specification

2.1 INTRODUCTION

The Geoscience Spatial Framework (GSF) is designed to act as a common data structure that can be populated by any 3D modelling package, and against which a range of visualisation applications tailored to particular end-users can be built. These applications will be built over time to meet evolving requirements. However in order to test that the GSF system works the DGSM Data Architecture sub-project will design and build a trial visualisation application against the GSF database. This document is designed to provide the specification for this GSF trial output application.

The DGSM models will be used both internally to BGS, to assist in developing added-value products and to aid scientific understanding, and will be made available directly to external users principally over the web. The GLOS will be the main vehicle for fulfilling the first of these roles, as the GLOS will retain the full richness of the original models. The GLOS will also be capable of producing VRML and other outputs for public dissemination, but the user will only have limited manipulation and customisation control over these. The GSF will therefore be the main means of fulfilling the external dissemination function and it must be possible for external users to access the GSF over the web on their own PCs. The GSF trial application therefore should run using a browser with no plug-ins or add-ons that are not freely downloadable.

The GSF trial application is designed to test the main elements that will be required in GSF output applications generally, while at the same time providing a useful deliverable in its own right. The main components of any GSF output application will be: retrieval from the GSF database against a range of spatial, metadata, and attribute criteria; the creation of a model structure from the GSF points that is consistent with the original; visualisation of the recreated model; dissemination of the model over the web; user interaction with the disseminated model to view it within a range of different pre-determined visualisations and to interrogate model elements for their attribution.

The GSF trial application will incorporate these components as part of a general ‘model visualisation’ application – in effect a 3D DIGMAPGB that is not tailored for any one specific user group. The techniques developed in the trial application can subsequently be used to build a range of more specific applications aimed at addressing specific problems or user groups.

The overall schema of the trial application is shown in figure 1.

2.2 SELECTION MODULE

The selection of data held in the GSF for display is the first process in the application and is illustrated in figure 2. There are three separate data selection processes that can be carried out:

- Spatial
- Metadata
- Attributes

The selection process must always start with spatial selection, and this is therefore the sole entry point to the application. Spatial selection can be used either independently or in combination with the other selection modes.

The application should only retrieve GSF points where the ‘Access Constraint’ field in the DGSM model metadata indicates that the points are available for public dissemination.

2.2.1 Spatial selection

The spatial selection functionality available should be similar to that in the Internet GDI, with an initial screen showing the whole UK and surrounding seas. From there it should be possible to zoom in to a chosen area either in standard steps centred on a user-defined point or by drawing a rectangle. There should also be the ability to zoom to areas centred on a place identified in a gazetteer (as in the ‘Desktop’ GDI).

Once the display area has been determined the user should be prompted to define the GSF data selection using one of the following options:

- Visible extent
- Drawn or keyed rectangle
- Drawn or keyed circle

The GSF spatial data selection module should build on work already done for the DGSM Spatial Gateway and DGSM Data Portal.

It should be possible to redo the spatial selection any number of times prior to moving on to the next stage.

On completion of spatial data selection the user should have the option of either refining the selection further by using metadata or attribute selection, or leaving the selection module and proceeding to list the titles of the models that have been selected as the first step to model display. The model titles apply to complete models as loaded to the GSF whereas the spatial selection may have only selected GSF points for part of a model. The list of models should flag those where only part of a model has been selected.

2.2.2 Metadata selection

The metadata selection module should work against the DGSM ‘Model’ Metadata as defined in Laxton et al (2001). The model metadata applies to complete models as loaded to the GSF, so metadata selection will select all GSF points for a model in the selection set that meet the

metadata selection criteria. Metadata selection must follow from spatial selection and should operate against a selection set comprising the subset of GSF points derived by spatial selection.

The user should be able to select using any or all of the following DGSM model metadata fields:

- Keyword
- Areal Extent
- Vertical Extent
- Bounding depth units
- Highest point
- Lowest point

Keyword selection should be assisted by access to the hierarchical thesaurus from which they are derived, using an interface similar to that developed for the DGSM metadata input application. Similarly Areal Extent selection should be assisted using an hierarchical listing like that used in the DGSM metadata input application.

Vertical Extent should be defined by selecting one or more values from the appropriate dictionary.

The user should also be able to specify a highest and/or lowest depth for selection, using units selected from a list of available 'Bounding depth units'. The application will not carry out any unit conversions, simply retrieving against the values held in the metadata, and so will fail to retrieve GSF points with depth ranges recorded in units different from those selected for retrieval. The user should be made aware of this possibility.

It should be possible to redo the metadata selection any number of times prior to moving on to the next stage.

On completion of metadata data selection the user should have the option of either refining the selection further by using attribute selection, or leaving the selection module and proceeding to list the titles of the models that have been selected as the first step to model display. The model titles apply to complete models as loaded to the GSF whereas the preceding spatial selection may have only selected GSF points for part of a model. The list of models should flag those where only part of a model has been selected.

2.2.3 Attribute selection

All the points in the GSF will be linked to one or more attribute tables and attribute selection provides a means of retrieving from the GSF against attribute criteria defined by the user. In particular this option provides a means of retrieving points of specific aspect (on the surface, above the surface, below the surface) whereas spatial and metadata selection will retrieve points of all aspects. Attribute selection must follow from spatial and/or metadata selection and should operate against a selection set comprising the subset of GSF points derived by the preceding selections.

There should be two routes into the attribute selection module, either directly from the spatial or metadata modules, or indirectly through selection from a list of models that has been generated as a result of spatial and/or metadata selection. This latter should be one of the options that can be selected from the list of models produced by the spatial and metadata selection modules (see section 3.1). The direct route will be useful for queries such as 'Find all reverse faults in area X' whereas the indirect route would enable queries such as 'Find the Lower and Middle Chalk from the 2002 SE England Chalk Model'.

Attribute selection should be based around an SQL query builder, as the range of different possible models and features will make it impractical to provide 'pre-cooked' attribute retrievals that are sensible for all selections. The query builder should first prompt the user with a list of those features, as defined in the DGSM data model, that exist within the GSF selection set that the user has generated from the preceding selections. One of these should be selected for query building. Once a feature is selected the user should be presented with a list of attributes for that feature, SQL operators, and possible attribute values derived from constraint dictionaries where appropriate, and guided through the construction of an SQL query.

The results of the query should be written to a selection set, the user informed if the query failed to select any points, and the user given the choice of proceeding to the display modules, adding to the query, or deleting the query. If the user chooses to add to the query then s/he is returned to the SQL query builder where another query can be built and its output added to the GSF points already in the selection set. If the user chooses to delete the query then the points selected by the query are removed from the selection set. This process can be continued until the user chooses to proceed to the display module.

2.3 OUTPUT AND DISPLAY COMPONENTS

2.3.1 List of models by title

The output from both the spatial and metadata selection modules is a list of the titles, derived from the model metadata, of the models that meet the selection criteria. It should be remembered that in the case of spatial selection only part of a model need lie within the area specified. At this point the user should have six options:

1. Return to the spatial selection module and start again (section 2.2.1)
2. Return to the metadata selection module and make a new metadata selection using the same spatial selection (section 2.2.2)
3. Select one or more models from the list for input to the attribute selection module (section 2.2.3)
4. Select a model to display its full DGSM model metadata (section 2.3.2)
5. Select one or more models for display (section 2.3.3)
6. Exit the application

2.3.2 Model metadata display

The DGSM model metadata provides information, mainly in free text fields, that can help a user decide if they wish to select a particular model for display. The model metadata display module should enable users to select one of the models from the list of models produced by the selection module, and to display all its model metadata fields. Users should be able to repeat this process as often as they wish with different models from the list.

2.3.3 Model display

Once the user is satisfied with the list of models selected one or more of the models can be selected for display. This should be achieved by simply clicking a selection box beside each model title in the list. Where attribute selection has been used this step is missed out and the user goes straight to the selection of display options, because the user has already been given the option of selecting specific models. In either case the end result is a set of GSF points to be displayed.

It will be necessary at this stage to ensure that the number of points selected is not excessive for display. If it is, the user should be given the choice of one of two options:

1. Return to the selection of listed models in order to select a different model from the list for display (this option will be invalid if attribute selection was used and should be blocked in that case) (section 2.3.1)
2. Return to the spatial selection component of the selection module to start the selection process again from the beginning (section 2.2.1)

The user should then be offered a list of display and export options which should include:

- Mesh/skeletal surface
- Rendered surface
- Block model
- Point cloud
- Export formats such as VRML and XMMML

The user should be allowed to input a vertical exaggeration if required.

Not all display options will be appropriate for all selection sets, for example a block model would not be appropriate if only one surface was selected. Ideally the application would present the user with just those display options appropriate for the selection set.

The application should then automatically generate the chosen display, carrying out the triangulation needed transparently, and using default colours to distinguish different surfaces and volumes.

The user should be able to interact with the resulting display with the following functions:

- turning the model to view from different angles
- zooming
- panning
- the ability to slice the model to view cross-sections

Once a model has been displayed the user should have five options:

1. Return to the selection of display/export options to display or export the same model using different display/export options (section 2.3.3)
2. Return to the selection of listed models in order to select a different model from the list for display (this option will be invalid if attribute selection was used and should be blocked in that case) (section 2.3.1)
3. Return to the spatial selection component of the selection module to start the selection process again from the beginning (section 2.2.1)
4. Enter the element attribute display module (section 2.3.4)
5. Exit the application

2.3.4 GSF point attribute display

All the points in the GSF will be linked to one or more attribute tables and attribute display provides a means for the user to view these attributes interactively. The user should be able to enter query mode when displaying the model and click on any point in the model to be informed of the feature that the point belongs to and then to be shown a listing of the attributes of the feature at that point. Where a point is part of more than one feature these should be listed and the

user asked to select one for attribute listing. Users should be able to repeat this process as often as they wish with different points in the model.

3 DGSM Metadata Application Specification for Selection, Retrieval and Evaluation

3.1 INTRODUCTION

The DGSM metadata (Laxton et al, 2001) is divided into three distinct types each designed to answer a particular type of question:

1. Data Discovery metadata is designed to tell modellers what suitable data exist for creating a particular model.
2. Model Discovery metadata is designed to tell model users what suitable models exist for a particular application.
3. Inference metadata is designed to tell model developers and advanced users how a model was developed.

This report provides the specification for the application that will enable the metadata to be used to address these questions. The specification is for an application to be used internally by BGS staff, although a similar application for external use could be readily adapted from it.

The report will describe separately that part of the application which uses each type of metadata, but they are nevertheless connected. A user evaluating data sub-sets may wish to view metadata describing how these were used to create models, as well as metadata describing the resulting models. Similarly a user wishing to investigate what models exist may wish to investigate how these were created and the data used to create them. The first step in the application will therefore be to present the user with the choice of running the application from a dataset investigation or a model investigation perspective.

The Inference Metadata Component of the application links a particular data sub-set to a particular model and must be approached from either the data sub-set or model perspective. Figure 3 provides an overview of the entire metadata application.

3.2 DATA DISCOVERY COMPONENT

The evaluation, retrieval and creation of DGSM data sub-sets is done using the DGSM 'Data Discovery Metadata' as defined in Laxton et al (2001). The purpose of this part of the DGSM metadata application is to allow a user to determine what data sets or sub-sets exist within a particular area, pertaining to a particular topic; to be able to access the metadata describing those data sets or sub-sets in order to evaluate them for use in modelling; and finally to either retrieve specific data sub-sets or to create new ones. The application can be used to evaluate data sets and sub-sets without creating new ones, and use of this application will not be the only means by which DGSM data sub-sets can be created. Figures 4 and 5 give the schema for the data discovery component of the application. This is divided into three modules, selection, evaluation and retrieval, each of which will be described separately.

3.2.1 Selection module

The selection of data sets or data sub-sets is the first part of the process. There are two separate data selection processes that can be carried out:

- Spatial

4. Metadata

The selection process must always start with spatial selection, and this is therefore the sole entry point to this part of the application. Spatial selection can be used either independently or in combination with metadata selection.

3.2.1.1 SPATIAL SELECTION

The spatial selection module should build on work already done for the DGSM Spatial Gateway and DGSM Data Portal.

The spatial selection functionality available should be similar to that in the Internet GDI, with an initial screen showing the whole UK and surrounding seas. From there it should be possible to zoom in to a chosen area either in standard steps centred on a user-defined point or by drawing a rectangle. There should also be the ability to zoom to areas centred on a place identified in a gazetteer (as in the 'Desktop' GDI).

Once the display area has been determined the user should be prompted to define data selection using one of the following options:

- Visible extent
- Drawn or keyed rectangle
- Drawn or keyed circle
- Drawn or keyed polygon

It should be possible to redo the spatial selection any number of times prior to moving on to the next stage.

On completion of spatial selection the user should have the option of either refining the selection further by using metadata selection, or leaving the selection module and proceeding to list the titles of the data sets and sub-sets that have been selected. The data sets and sub-sets selected may be only partially included within the selection area and where this is the case they should be flagged in the list.

3.2.1.2 METADATA SELECTION

The metadata selection module should use the DGSM Data Discovery Metadata. Metadata selection must follow from spatial selection and should operate against a selection set comprising the subset of data sets and sub-sets derived by spatial selection.

The user should be able to select using any or all of the following DGSM discovery metadata fields:

- Keyword
- Areal Extent
- Vertical Extent
- Bounding depth units
- Highest point
- Lowest point

Keyword selection should be assisted by access to the hierarchical thesaurus from which they are derived, using an interface similar to that developed for the DGSM metadata input application. Similarly Areal Extent selection should be assisted using an hierarchical listing like that used in the DGSM metadata input application.

Vertical Extent should be defined by selecting one or more values from the appropriate dictionary.

The user should also be able to specify a highest and/or lowest depth for selection, using units selected from a list of available 'Bounding depth units'. The application should carry out certain common unit conversions (eg feet to metres), but will not be able to cope with all possible conversions. In addition some conversions are not possible (eg metres to seconds of two way time). The application should list data sets or sub-sets not retrieved because unit conversion has not been carried out.

It should be possible to redo or refine the metadata selection any number of times prior to moving on to the next stage.

On completion of metadata selection the user proceeds to list the titles of the data sets and sub-sets that have been selected. The data sets and sub-sets selected may be only partially included within the selection area and where this is the case it should be flagged in the list.

3.2.2 Evaluation module

3.2.2.1 LIST OF DATA SETS AND SUB-SETS

The output from both spatial and metadata selection is a list of the titles, derived from the data discovery metadata, of the data sets and sub-sets that meet the selection criteria. It should be remembered that only part of a data set or sub-set need lie within the area specified. At this point the user should have six options:

1. Return to the spatial selection module and start again (section 3.2.1.1)
2. Return to the metadata selection module and make a new metadata selection using the same spatial selection (section 3.2.1.2)
3. Select a data sub-set to display a list of models that it has been used to produce (section 3.2.2.3)
4. Select a data set or sub-set to display full DGSM data discovery metadata (section 3.2.2.2)
5. Obtain a data sub-set (section 3.2.3)
6. Exit the application

3.2.2.2 DATA DISCOVERY METADATA DISPLAY

If the user chooses to display full DGSM data discovery metadata then s/he should be offered a further three options:

1. Display the full DGSM data discovery metadata of the data set or sub-set
2. Display the full DGSM data discovery metadata of the parent of the data set or sub-set selected
3. Display the full DGSM data discovery metadata of any child data sub-sets of the data set or sub-set selected

The DGSM data discovery metadata provides information, mainly in free text fields, that can help a user decide if they wish to select a particular data set or subset. The data discovery

metadata display component should enable users to select one of the data sets or sub-sets from the list (section 3.2.2.1), and to display, print, or save to file **all** its data discovery metadata fields. Users should be able to repeat this process as often as they wish with different data sets or sub-sets from the list.

Some fields in the BGS-geoIDS discovery metadata are not included in the DGSM data discovery metadata because it is assumed that the data sub-set will inherit these from the BGS-geoIDS data set at the top of the hierarchy. These inherited metadata fields should be displayed by the application along with those specific to the data sub-set.

The application should provide the user with the option to see the data discovery metadata for the parent of a data sub-set selected from the list. This parent could be either the BGS-geoIDS data set at the top of the hierarchy or an intermediate data sub-set. The user should have the option to add the parent to the list (section 3.2.2.1) of data sets and sub-sets previously produced.

The application should provide the user with the option to see the data discovery metadata for any child data sets produced from the data set or data sub-set selected from the list. The user should have the option to add any child to the list (section 3.2.2.1) of data sets and sub-sets previously produced.

3.2.2.3 LIST OF MODELS PRODUCED FROM A DATA SUB-SET

A data sub-set will have been used to produce at least one model, but may have been used to produce more. The application should allow the user to select a data sub-set from the list (section 3.2.2.1) and to generate for this data sub-set a list of the models that the data sub-set has been used to produce. It should be remembered that the spatial extent of the model may not be the same as that of the data sub-set. At this point the user should have five options:

1. Add to the list of models (section 3.2.2.4)
2. Select a model to display its full DGSM model discovery metadata (section 3.2.2.5)
3. Select a model to review its inference metadata (section 3.4)
4. Select a model for display or retrieval (section 3.3.3)
5. Return to the list of data sets and sub-sets obtained from the selection module (section 3.2.2.1)

3.2.2.4 ADD PARENT OR CHILD MODEL TO THE LIST OF MODELS

The DGSM model discovery metadata contains a field to relate component models to higher order models of which they are a component (the parent). This is a mechanism whereby a hierarchy of model components can be established, for example linking individual surface models to a block model of which they form part or linking individual block models to form a regional model.

The application should allow a model to be selected from the list (section 3.2.2.3) and the title of the model one order up in the hierarchy (the parent), where one exists, to be added to the list of models. This procedure can then be repeated with that model enabling the entire upward hierarchy to be added to the list. Once added to the list of models the application should allow these models higher in the hierarchy to be handled in exactly the same way as the lower order models.

The application should also allow a model to be selected from the list (section 3.2.2.3) and the title of any model one order down in the hierarchy (a child), where one exists, to be added to the list of models. This procedure can then be repeated with that model enabling the entire downward hierarchy to be added to the list. Once added to the list of models the application

should allow these models lower in the hierarchy to be handled in exactly the same way as the higher order models.

3.2.2.5 MODEL DISCOVERY METADATA DISPLAY

The DGSM model discovery metadata provides information, mainly in free text fields, about a particular model. This information may help a user get a better understanding of the purposes for which a data sub-set was created and used, and thus whether the data sub-set will be useful for the user's current requirement. The model metadata display module should enable users to select one of the models from the list of models produced from a data sub-set (section 3.2.2.3), and to display, print, or save to file **all** its model discovery metadata fields. Users should be able to repeat this process as often as they wish with different models from the list.

3.2.3 Data retrieval module

The data retrieval module is designed to allow the user to proceed from data set and sub-set evaluation to the retrieval of data based on the results of the evaluation. This can either be the retrieval of an existing data sub-set or the creation of a new data sub-set. The use of this module is optional – the user can terminate the application after data evaluation. Note that retrieval of data using this application is not the only means by which a data sub-set for modelling can be created. Figure 5 gives the schema for the data retrieval module.

All data retrieval has, as a starting point, the selection of a data set or sub-set for retrieval from the list of data sets and sub-sets used in either the Data Discovery evaluation module (see section 3.2.2.1) or the Model Discovery evaluation module (section 3.3.2.4). From here the user should be given two options:

1. Retrieve the data sub-set (section 3.2.3.1)
2. Define a new data sub-set (section 3.2.3.2)

3.2.3.1 RETRIEVE AN EXISTING DATA SUB-SET

This option is only available if a data sub-set, rather than a data set, has been selected for retrieval. The user is prompted with a list of the fields, as defined in the parent BGS-geoIDS data set, and asked to select those which should be written to an Ascii output file. At least x, y and z fields must be selected. The user should specify the output file directory and format. The format should be selected from a list including tab delimited, comma delimited, XML and others suitable for input to BGS modelling software packages.

On completion the user is returned to the evaluation module from which the data retrieval module was entered and the list of data sets and sub-sets previously selected (section 3.2.2.1 or 3.3.2.4).

3.2.3.2 DEFINE SPATIAL RETRIEVAL CRITERIA FOR A NEW DATA SUB-SET

The user should have the choice of either creating a new data sub-set for the area previously defined by spatial selection (see section 3.2.1.1 or 3.3.1.1), or creating a new data sub-set using new spatial selection criteria. For example for data evaluation a user may have defined an area graphically, but for data retrieval will require a more precise keyed input of coordinates.

If a user chooses to define new spatial selection criteria then the spatial selection functionality available should be the same as that in the selection modules (section 3.2.1.1 and 3.3.1.1), with an initial screen showing the whole UK and surrounding seas. From there it should be possible to zoom in to a chosen area either in standard steps centred on a user-defined point or by drawing a rectangle. There should also be the ability to zoom to areas centred on a place identified in a gazetteer (as in the 'Desktop' GDI).

Once the display area has been determined the user should be prompted to define a data retrieval area using one of the following options:

- Visible extent
- Drawn or keyed rectangle
- Drawn or keyed circle
- Drawn or keyed polygon

The spatial criteria defined should be used to retrieve an initial selection set. This can subsequently be refined using non-spatial SQL-based retrieval if required.

3.2.3.3 NON-SPATIAL REFINEMENT OPTIONS

On completion of spatial retrieval the user should be presented with five options:

1. Proceed without non-spatial refinement (section 3.2.3.6)
2. Define non-spatial refinement using new SQL (section 3.2.3.4)
3. Define non-spatial refinement using existing SQL from another data sub-set of the same parent (section 3.2.3.5)
4. Redefine the spatial retrieval criteria for the same data set or sub-set (section 3.2.3.2)
5. Return to the evaluation module from which the data retrieval module was entered and the list of data sets and sub-sets previously selected (section 3.2.2.1 or 3.3.2.4)

If no non-spatial refinement is required then a list of the data identifiers of all records retrieved by the spatial selection should be produced (section 3.2.3.6)

3.2.3.4 DEFINE NON-SPATIAL REFINEMENT WITH NEW SQL

The user should be presented with a window to allow an SQL query to be built for the data set or sub-set selected. This should allow the selection of appropriate commands and operators as well as the field names from the selected data set or sub-set. Spatial data fields should be excluded. Where fields are constrained by a dictionary then it should be possible to select values from these for the appropriate field.

The SQL should be run against the initial selection set created by the earlier spatial retrieval (section 3.2.3.2), and should produce a list of the data identifiers of all records retrieved (section 3.2.3.6).

3.2.3.5 DEFINE NON-SPATIAL REFINEMENT WITH EXISTING SQL

This option allows the user to select a data set or sub-set and refine it using SQL stored in the 'Inclusion Criteria SQL' field of any data sub-set derived from the same BGS-geoIDS data set.

The application should first provide the user with a list of all those data sub-sets derived from the same BGS-geoIDS data set. The user can then select one of these data sub-sets to retrieve the stored SQL and run it against the initial selection set created by the earlier spatial retrieval (section 3.2.3.2), to produce a list of the data identifiers of all records retrieved (section 3.2.3.6).

In order to assist the user to select the appropriate SQL the application should provide an SQL viewing function. This should allow the user to select a data sub-set from the list of all those data sub-sets derived from the same BGS-geoIDS data set and then view the contents of the 'Inclusion Criteria SQL' field for that data sub-set.

As database definitions change the use of the SQL stored in the metadata may fail. If this happens the user should be given the option of either returning to the list of non-spatial

refinement options (section 3.2.3.3) or proceeding directly to the definition of non-spatial refinement using new SQL (section 3.2.3.4). If this latter option is chosen the SQL which failed should be displayed in a window in addition to the SQL definition window, from where it can be used as a guide or pasted into the new SQL. The database error message should also be provided and the user given the option to e-mail it to the relevant application or database manager.

3.2.3.6 LIST OF ALL DATA INSTANCES RETRIEVED

All the options for creating a new data sub-set produce as an intermediate step a list of the data identifiers for the records in the selection set. At this point the user has three options:

1. Return to the list of non-spatial refinement options (section 3.2.3.3)
2. Proceed to include or exclude specific data instances (section 3.2.3.7)
3. Proceed to create a new data sub-set for the data instances in the selection set (section 3.2.3.8)

3.2.3.7 INCLUDE/EXCLUDE SPECIFIC DATA INSTANCES

In many cases the user may wish to exclude specific data instances that meet the general selection criteria, or to include specific data instances that fail to meet the general selection criteria. The application should allow individual records to be selected for exclusion from the selection set. The application should also allow the user to key in the data identifiers of data instances, not in the selection set, for inclusion.

The Data Discovery metadata of data sub-sets includes details of included and excluded data instances either in a set of individual fields (*Included data item identifier; Data item inclusion criteria; Reasons for data item inclusion criteria; Excluded data item identifier; Data item exclusion criteria; Reasons for data item exclusion criteria*) or in two fields pointing to a file or URL containing the same information (*Included data items file; Excluded data items file*). Where a data sub-set, as opposed to a data set, is being used to create the records in the selection set the application should give the user the option of viewing this metadata. The application should then allow the user to select all the data instances listed, or to select individual data instances from the list, for inclusion or exclusion. The data identifiers for these records are then added or removed from the selection set.

3.2.3.8 CREATE NEW DATA SUB-SET

The starting point for the creation of a new data sub-set is the list of records in the selection set. The user is then prompted with a list of appropriate fields, as defined in the BGS-geoIDS data set from which the selection set is derived, and asked to select which ones should be written to an Ascii output file. At least x, y and z fields must be selected. The user should specify the output file directory and format. The format should be selected from a list including tab delimited, comma delimited, XML and others suitable for input to BGS modelling software packages. The application should create a file containing the selected fields for all records identified in the selection set.

On completion the user is taken to the DGSM Metadata Entry Application (Holmes, 2002)) in order to enter Data Discovery metadata for the data sub-set just created. The user is then returned to the evaluation module from which the data retrieval module was entered, and the list of data sets and sub-sets previously selected (section 3.2.2.1 or 3.3.2.4).

3.3 MODEL DISCOVERY COMPONENT

The selection, evaluation, visualisation and retrieval of DGSM models is done using the DGSM 'Model Discovery Metadata' as defined in Laxton et al (2001). The purpose of this part of the

DGSM metadata application is to allow a user to determine what models exist within a particular area, pertaining to a particular topic; to enable access to the metadata describing those models in order to evaluate their suitability for a particular purpose; and finally to either visualise the model in the case of the GSF, or retrieve it for input to a modelling package or visualisation software in the case of the GLOS. This part of the application is similar to the 'GSF Trial Output Application Specification' described in section 2, but with the difference that it provides access to the GLOS as well as the GSF.

Figure 6 gives the schema for the model discovery component of the application. This is divided into three modules, selection, evaluation and display/retrieval, each of which will be described separately.

3.3.1 Selection module

The selection of models is the first process in the application and the schema for the selection component is illustrated in figure 7. There are three separate selection processes that can be carried out, although attribute selection is only available for the GSF:

- Spatial
- Metadata
- Attribute

The selection process must always start with spatial selection (section 3.3.1.1), and this is therefore the sole entry point to this part of the application. Spatial selection can be used either independently or in combination with the other selection modes.

3.3.1.1 SPATIAL SELECTION

The selection module should build on work already done for the DGSM Spatial Gateway and DGSM Data Portal.

The spatial selection functionality available should be similar to that in the Internet GDI, with an initial screen showing the whole UK and surrounding seas. From there it should be possible to zoom in to a chosen area either in standard steps centred on a user-defined point or by drawing a rectangle. There should also be the ability to zoom to areas centred on a place identified in a gazetteer (as in the 'Desktop' GDI).

Once the display area has been determined the user should be prompted to define the selection using one of the following options:

- Visible extent
- Drawn or keyed rectangle
- Drawn or keyed circle
- Drawn or keyed polygon

It should be possible to redo the spatial selection any number of times prior to moving on to the next stage.

On completion of spatial data selection the user should have the option of either refining the selection further by using metadata (section 3.3.1.2) or attribute selection (section 3.3.1.3), or leaving the selection module and proceeding to the evaluation module and the list of model titles that have been selected (section 3.3.2.1). The model titles apply to complete models whereas, in the case of the GSF, the spatial selection may have only selected points for part of a model. In the case of the GLOS complete GLOS objects should be selected if they are contained in, or overlapped by, the selection area. The list of GSF models should flag those where only part of a

model has been selected and the list of GLOS objects should flag those which overlap the selection area.

3.3.1.2 METADATA SELECTION

The metadata selection module should use the DGSM Model Metadata as defined in Laxton et al (2001). Metadata selection must follow from spatial selection and should operate against a selection set comprising the subset of GSF points or GLOS objects derived by spatial selection.

The user should be able to select using any or all of the following DGSM model metadata fields:

- Keyword
- Areal Extent
- Vertical Extent
- Bounding depth units
- Highest point
- Lowest point

Keyword selection should be assisted by access to the hierarchical thesaurus from which they are derived, using an interface similar to that developed for the DGSM metadata input application. Similarly Areal Extent selection should be assisted using an hierarchical listing like that used in the DGSM metadata input application.

Vertical Extent should be defined by selecting one or more values from the appropriate dictionary.

The user should also be able to specify a highest and/or lowest depth for selection, using units selected from a list of available 'Bounding depth units'. The application should carry out certain common unit conversions (eg feet to metres), but will not be able to cope with all possible conversions and in addition some conversions are not possible (eg metres to seconds of two way time). The application should list data sets or sub-sets not retrieved because unit conversion has not been carried out.

It should be possible to redo the metadata selection any number of times prior to moving on to the next stage.

On completion of metadata selection the user should have the option of either refining the selection further by using attribute selection (section 3.3.1.3), or leaving the selection module and proceeding to list the titles of the models that have been selected (section 3.3.2.1). The first option is only available for GSF points. The model titles apply to complete models whereas in the case of the GSF the spatial selection may have only selected points for part of a model. In the case of the GLOS complete GLOS objects will be selected if they are contained in, or overlapped by, the selection area. The list of GSF models should flag those where only part of a model has been selected and the list of GLOS objects should flag those which overlap the selection area.

3.3.1.3 ATTRIBUTE SELECTION

All the points in the GSF are linked to one or more attribute tables and attribute selection provides a means of retrieving from the GSF against attribute criteria defined by the user. In particular this option provides a means of retrieving points of specific aspect (on the surface, above the surface, below the surface) whereas spatial and metadata selection will retrieve points of all aspects. Attribute selection must follow from spatial and/or metadata selection and should operate against a selection set comprising the subset of GSF points derived by the preceding selections.

There should be two routes into the attribute selection module, either directly from the spatial or metadata modules, or indirectly through selection from a list of models that has been generated as a result of spatial and/or metadata selection. This latter should be one of the options that can be selected for the list of GSF and GLOS models produced by the spatial and metadata selection modules (see section 3.3.2.1), but will only be available for the GSF models on the list. The direct route will be useful for queries such as ‘Find all reverse faults in area X’ whereas the indirect route would enable queries such as ‘Find the Lower and Middle Chalk from the 2002 SE England Chalk Model’.

Attribute selection should be based around an SQL query builder, as the range of different possible models and features will make it impractical to provide ‘pre-cooked’ attribute retrievals that are sensible for all selections. The query builder should first prompt the user with a list of those features, as defined in the DGSM data model, that exist within the GSF selection set that the user has generated from the preceding selections. One of these should be selected for query building. Once a feature is selected the user should be presented with a list of attributes for that feature, SQL operators, and possible attribute values derived from constraint dictionaries where appropriate, and guided through the construction of an SQL query.

The results of the query should be written to a selection set, the user informed of the number of records found, and the user given the choice of proceeding to the display module (section 3.3.3), adding to the query, or deleting the query. If the user chooses to add to the query then s/he is returned to the SQL query builder where another query can be built and its output added to the GSF points already in the selection set. If the user chooses to delete the query then the results of the query are removed from the selection set and the user returned to the SQL query builder where another query can be built. This process can be continued until the user chooses to proceed to the display module.

3.3.2 Evaluation module

3.3.2.1 LIST OF MODELS BY TITLE

The output from both spatial and metadata selection is a list of the titles, derived from the model metadata, of the models that meet the selection criteria. The list should also clearly identify whether a model is in the GSF or is a GLOS object. In many cases both formats will exist for the same model, and where this is the case these should be listed separately. It should be remembered that in the case of spatial selection only part of a GSF model need lie within the area specified and a GLOS object may extend beyond the area specified. Only that part of a GSF model within the specified area should be retrieved. At this point the user should have eight options:

1. Return to the spatial selection module and start again (section 3.3.1.1)
2. Return to the metadata selection module and make a new metadata selection using the same spatial selection (section 3.3.1.2)
3. Select one or more GSF models from the list for input to the attribute selection module (section 3.3.1.3)
4. Select a model to add the title of the parent or child model to the list of models (section 3.3.2.2)
5. Select a model to display its full DGSM model discovery metadata (section 3.3.2.3)
6. Select a model to display a list of data sub-sets from which the model has been derived (section 3.3.2.4)
7. Select one or more models for display (section 3.3.3)
8. Exit the application

3.3.2.2 ADD PARENT OR CHILD MODEL TO THE LIST OF MODELS

The DGSM model discovery metadata contains a field to relate component models to higher order models of which they are a component (the parent). This is a mechanism whereby a hierarchy of model components can be established, for example linking individual surface models to a block model of which they form part or linking individual block models to form a regional model.

The application should allow a model to be selected from the list (section 3.3.2.1) and the title of the model one order up in the hierarchy, where one exists, to be added to the list of models. This procedure can then be repeated with that model enabling the entire hierarchy to be added to the list. Once added to the list of models the application should allow models higher in the hierarchy to be handled in exactly the same way as the lower order models.

The application should also allow a model to be selected from the list (section 3.3.2.1) and the title of any model one order down in the hierarchy (a child), where one exists, to be added to the list of models. This procedure can then be repeated with that model enabling the entire downward hierarchy to be added to the list. Once added to the list of models the application should allow these models lower in the hierarchy to be handled in exactly the same way as the higher order models.

3.3.2.3 MODEL DISCOVERY METADATA DISPLAY

The DGSM model discovery metadata provides information, mainly in free text fields, that can help a user decide if they wish to select a particular model. The model discovery metadata display module should enable users to select one of the models from the list (section 3.3.2.1), and to display, print, or save to file **all** its model discovery metadata fields. Users should be able to repeat this process as often as they wish with different models from the list.

3.3.2.4 LIST OF DATA SUB-SETS FROM WHICH THE MODEL HAS BEEN DERIVED

A model will be derived from at least one data sub-set but it may have been derived from several. The application should allow the user to select a model from the list (section 3.3.2.1) and to produce a list of the data sub-sets from which that model was derived. It should be remembered that the spatial extent of the model may not be the same as the data sub-set. At this point the user should have four options:

1. Select a data sub-set to display its full DGSM data discovery metadata (section 3.3.2.5)
2. Select a data sub-set to review its inference metadata (section 3.4)
3. Obtain a data sub-set (section 3.2.3)
4. Return to the list of models (section 3.3.2.1)

3.3.2.5 DATA DISCOVERY METADATA DISPLAY

The DGSM data discovery metadata provides information, mainly in free text fields, about a particular data sub-set. This information may help a user get a better understanding of what a model has been derived from, and thus whether the model will be useful for the user's current requirement. The data discovery metadata display module should enable users to select one of the data sub-sets from the list of data sub-sets from which a model has been derived (section 3.3.2.4), and to display, print, or save to file **all** its data discovery metadata fields. Users should be able to repeat this process as often as they wish with different data sub-sets from the list.

3.3.3 Display and retrieval module

The display and retrieval module is designed to allow the user to proceed from model evaluation to the display, in the case of the GSF, or retrieval in the case of the GLOS, of models based on the results of the evaluation. The use of this module is optional – the user can terminate the application after model evaluation.

All model display and retrieval other than where attribute selection has been used has, as a starting point, the selection of a model from the list of models used in either the model evaluation module (see section 3.3.2.1) or the data evaluation module (section 3.2.2.3). Where attribute selection has been used the user goes straight to the GSF display module without the need to select a model from the list (section 3.3.1.3).

Once the user is satisfied with the list of models selected one or more models can be chosen for display or retrieval. This should be achieved by simply clicking a selection box beside each model title in the list. The options available will depend on whether the user chooses a GSF or GLOS model. If a GLOS object is selected then the user will be taken to the GLOS object retrieval module (section 3.3.3.3) and if a GSF model is selected the user will be taken to the GSF model display module (section 3.3.3.1).

3.3.3.1 GSF MODEL DISPLAY

The first step is to ensure that the number of GSF points selected is not excessive for display. If it is, the user should be given the choice of one of two options:

1. Return to the evaluation module from which the display module was entered, and the list of models previously selected (section 3.3.2.1 or 3.2.2.3) (this option will be invalid if attribute selection was used and should be blocked in that case)
2. Return to the spatial selection module previously used to define the extent (section 3.3.1.1 or 3.2.1.1) to start the selection process again from the beginning

If the number of GSF points is not excessive the user should then be offered a list of display and export options which should include:

- Mesh/skeletal surface
- Rendered surface
- Block model
- Point cloud
- Export formats such as VRML and XMMML

The user should be allowed to input a vertical exaggeration if required.

Not all display options will be appropriate for all selection sets, for example a block model would not be appropriate if a model containing only one surface was selected. Ideally the application would present the user with just those display options appropriate for the selection set.

The application should then automatically generate the chosen display, carrying out the triangulation needed transparently, and using default colours to distinguish different surfaces and volumes.

The user should be able to interact with the resulting display with the following functions:

- turning the model to view from different angles
- zooming
- panning

- the ability to slice the model to view cross-sections

Once a model has been displayed the user should have three options:

1. Return to the selection of display/export options to display or export the same model using different display/export options (section 3.3.3.1)
2. Return to the evaluation module from which the display module was entered, and the list of models previously selected (section 3.3.2.1 or 3.2.2.3) (this option will be invalid if attribute selection was used and should be blocked in that case)
3. Enter the GSF point attribute display module (section 3.3.3.2)

3.3.3.2 GSF POINT ATTRIBUTE DISPLAY

All the points in the GSF are linked to one or more attribute tables and attribute display provides a means for the user to view these attributes interactively. The user should be able to enter query mode when displaying the model and click on any GSF point in the model to be informed of the feature that the point belongs to and then to be shown a listing of the attribute values of the feature at that point. Where a point is part of more than one feature these should be listed and the user asked to select one for attribute listing. Users should be able to repeat this process as often as they wish with different points in the model.

3.3.3.3 GLOS OBJECT RETRIEVAL

GLOS model objects include both files that can only be used within the modelling environment in which they were generated, and export files and visualisations such as VRMLs. In order to use a GLOS object it is necessary to know what type of object it is and this information is held in the GLOS Technical Metadata. As a first step in retrieval the application should therefore present the user with a listing of the relevant GLOS technical metadata about the object selected. A GLOS model may comprise several files and the technical metadata will list these.

At this point the user has two options:

1. Return to the evaluation module from which the display module was entered, and the list of models previously selected (section 3.3.2.1 or 3.2.2.3)
2. Proceed to retrieve the GLOS object

If the user chooses to proceed and retrieve the GLOS object s/he should be prompted to specify a directory to receive the files and should have the option of renaming them if required. However as renaming may cause problems in some circumstances a warning should be given if this option is exercised. On completion the user should be returned to the evaluation module from which the display module was entered, and the list of models previously selected (section 3.3.2.1 or 3.2.2.3).

3.4 INFERENCE METADATA COMPONENT

A user may wish to understand more about how a data sub-set was used to create a particular model. The application should allow the user to select a model from the list of models produced from a data sub-set (section 3.2.2.3), or to select a data sub-set from the list of data sub-sets from which a model has been derived (section 3.3.2.4), and to review the DGSM inference metadata for this data sub-set and model combination.

The DGSM inference metadata, as defined in Laxton et al (2001), provides information on how a particular data sub-set was used in the creation of a particular model. It has no meaning outside of the context of a particular data sub-set and model combination, and therefore the Inference Metadata component can only be accessed through either the Model Discovery or Data Discovery components of the application (Figure 3).

The purpose of the inference metadata is to allow a user to understand better what a model was based on, not just in terms of what data was used in its creation, but also about how that data was used and the reasoning that went into the creation of the model. This understanding will help a user in selecting a model and in its subsequent use. Approached from the data sub-set direction the inference metadata provides an understanding of how a data sub-set has been used in previous modelling exercises, and thus provides guidance on how it or similar data sub-sets might be used in the future.

The inference metadata comprises, in the main, a set of structured free text fields, and the inference metadata component of the application simply provides a listing of these for the model and data sub-set combination that has previously been selected in either the Model Discovery or Data Discovery components of the application. The user is returned to the application component from which the inference metadata component was entered and from where other model and data sub-set combinations can be selected for inference metadata review. The application does not allow a 'cross-over' from either the Model Discovery component through the Inference Metadata component to the Data Discovery component or vice versa as this would disrupt the sequence of selection and evaluation that led to the list of models and data sub-sets used to determine the inference metadata to review.

References

LAXTON, J. L., HOLMES, K. A., MCKENZIE, A. A. 2001. Metadata Specification for the Digital Geoscience Spatial Model (DGSM). *British Geological Survey Internal Report*, IR/01/61.

HOLMES, K. A. 2002. Metadata User Guide. *Unpublished*.

Figure 1: Overall application schema

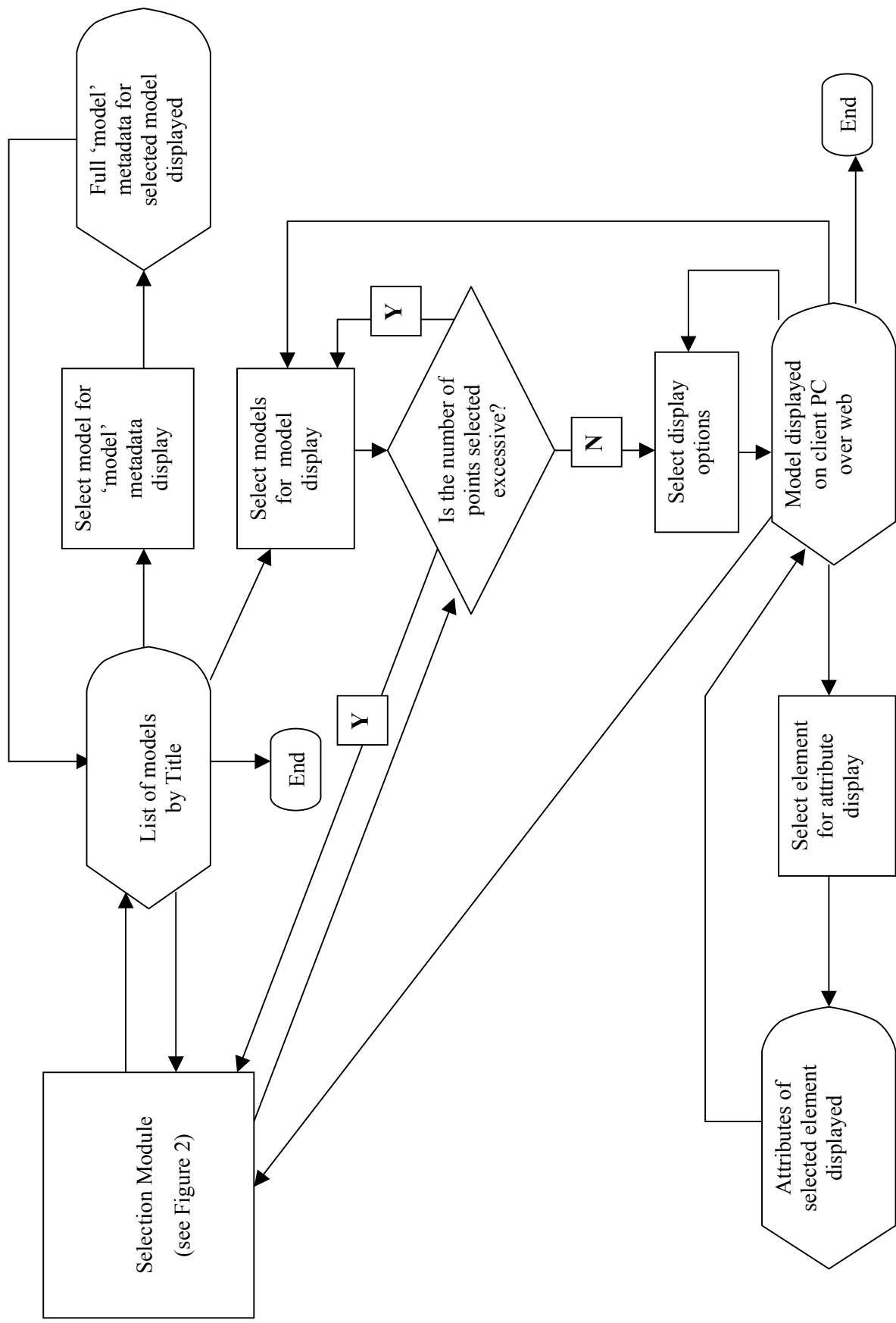


Figure 2: Selection module schema

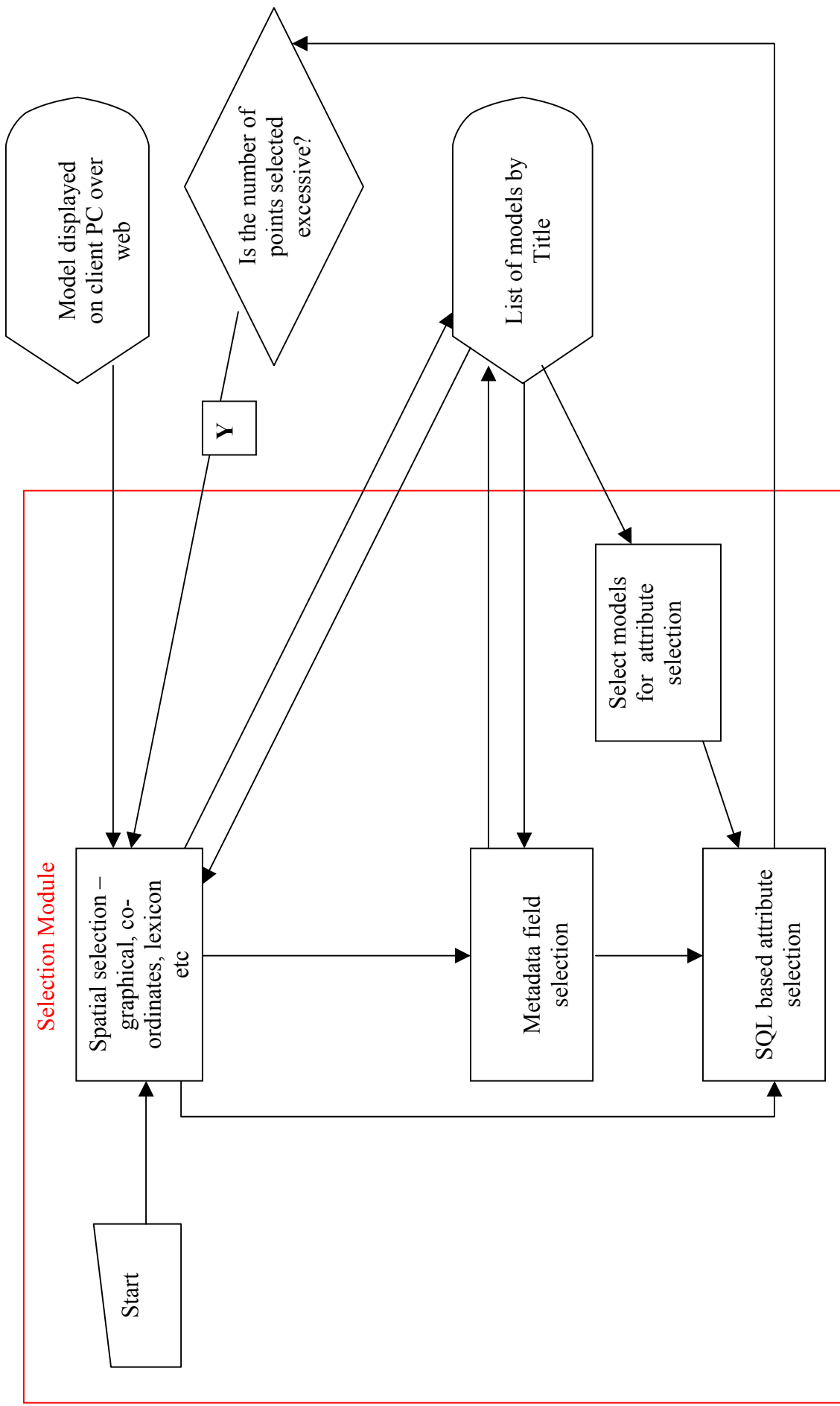


Figure 3: Overview of DGSM metadata application

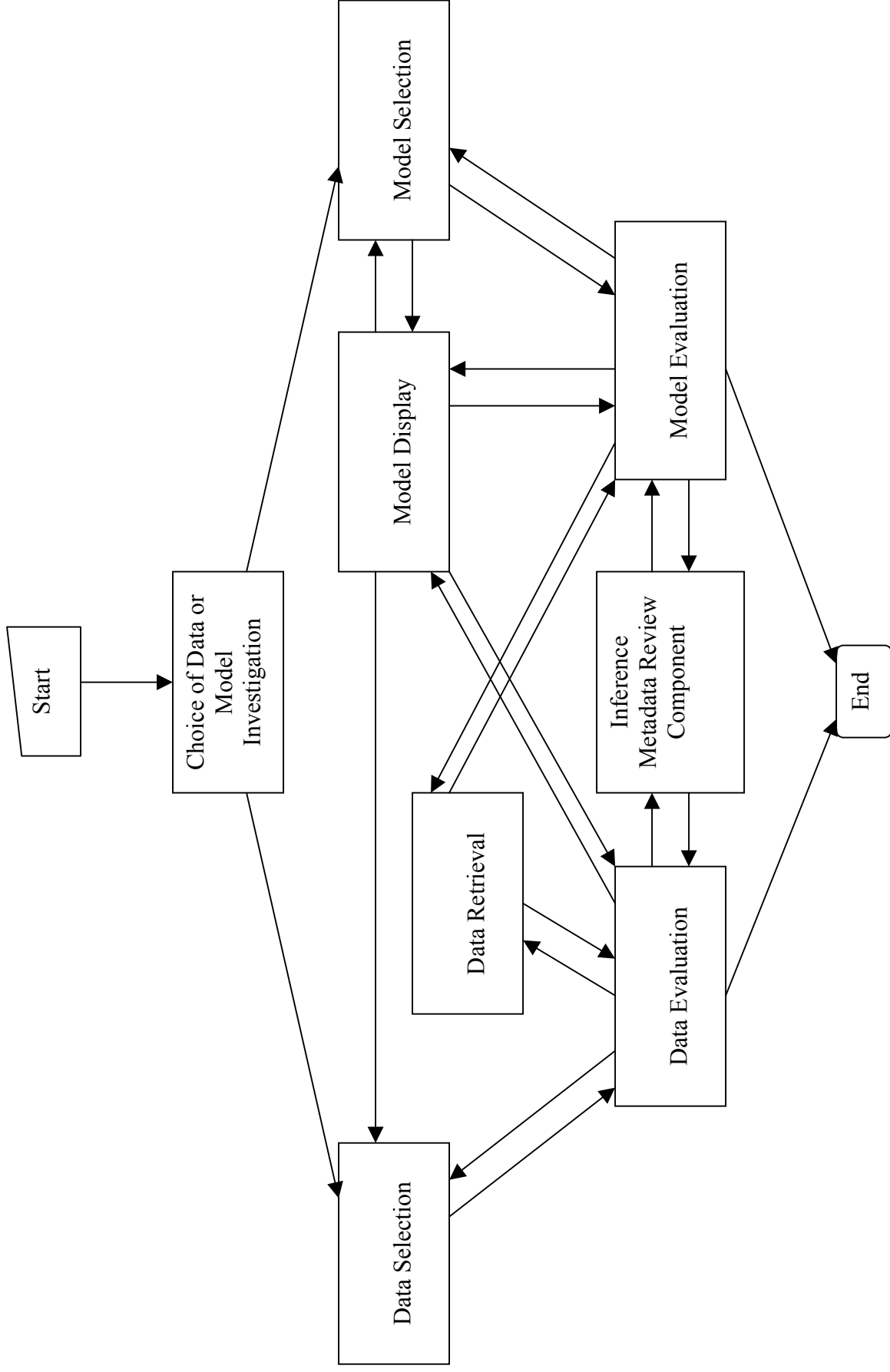


Figure 4: Data discovery component application schema

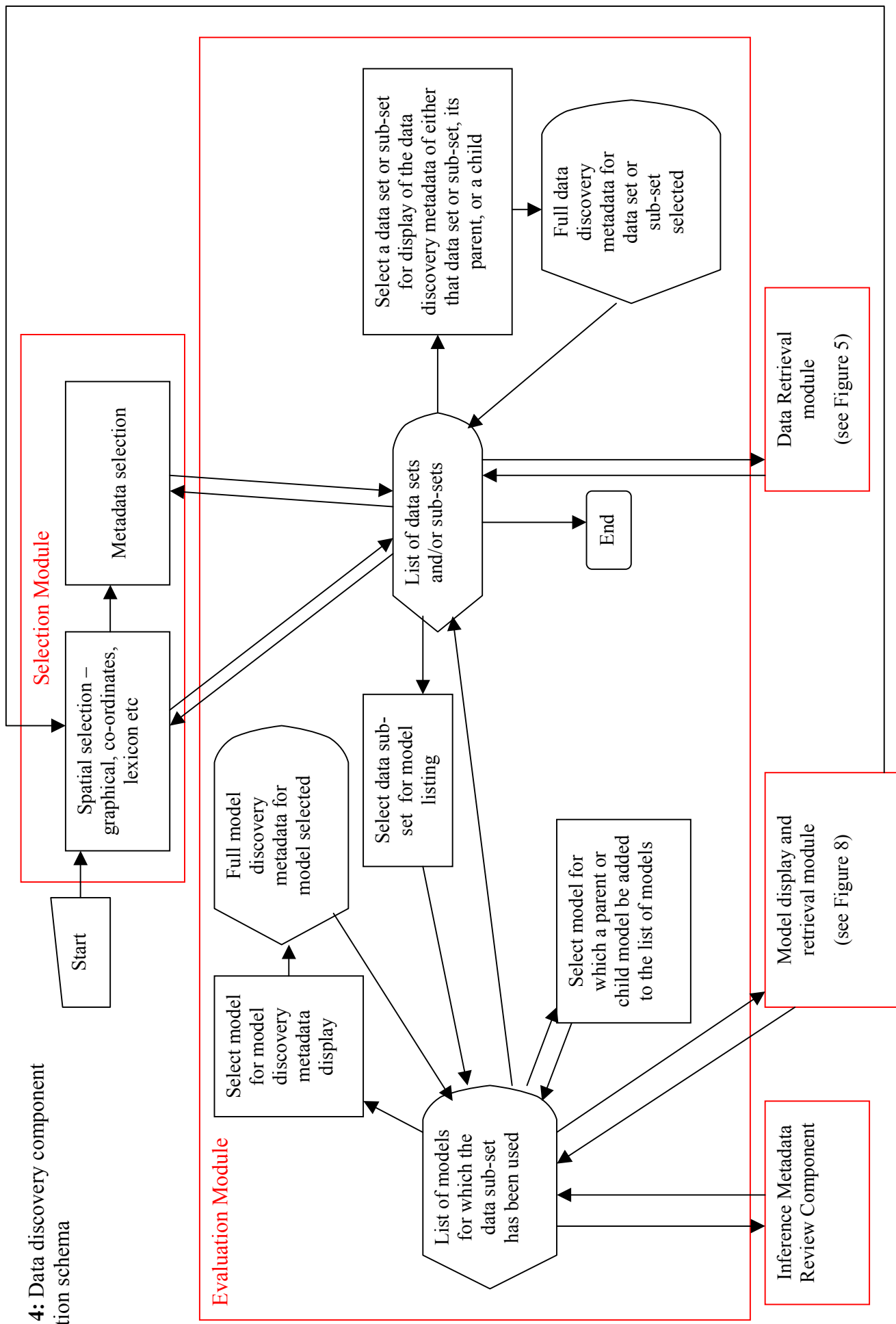


Figure 5: Data retrieval module schema

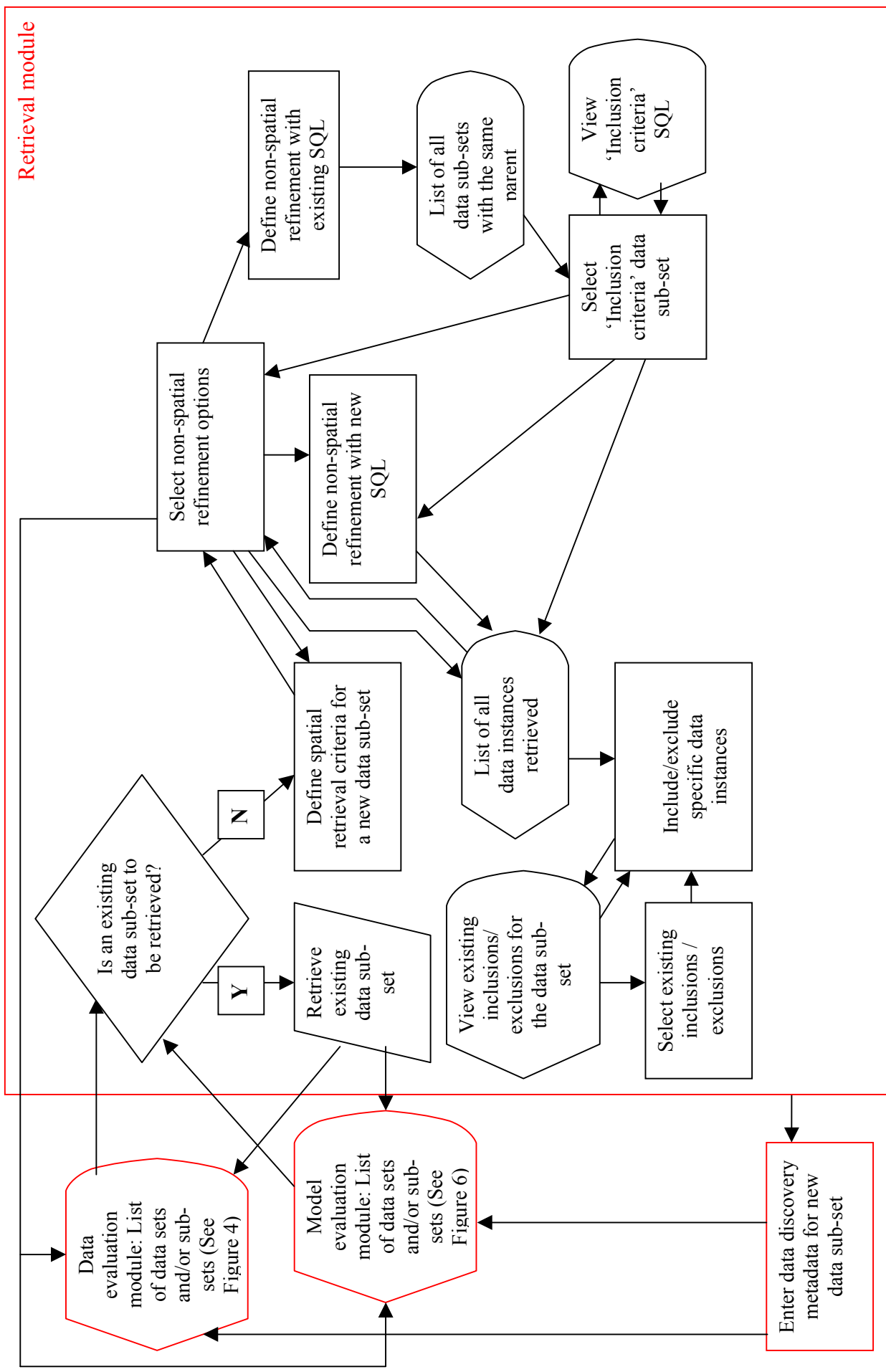


Figure 6: Model discovery component application schema

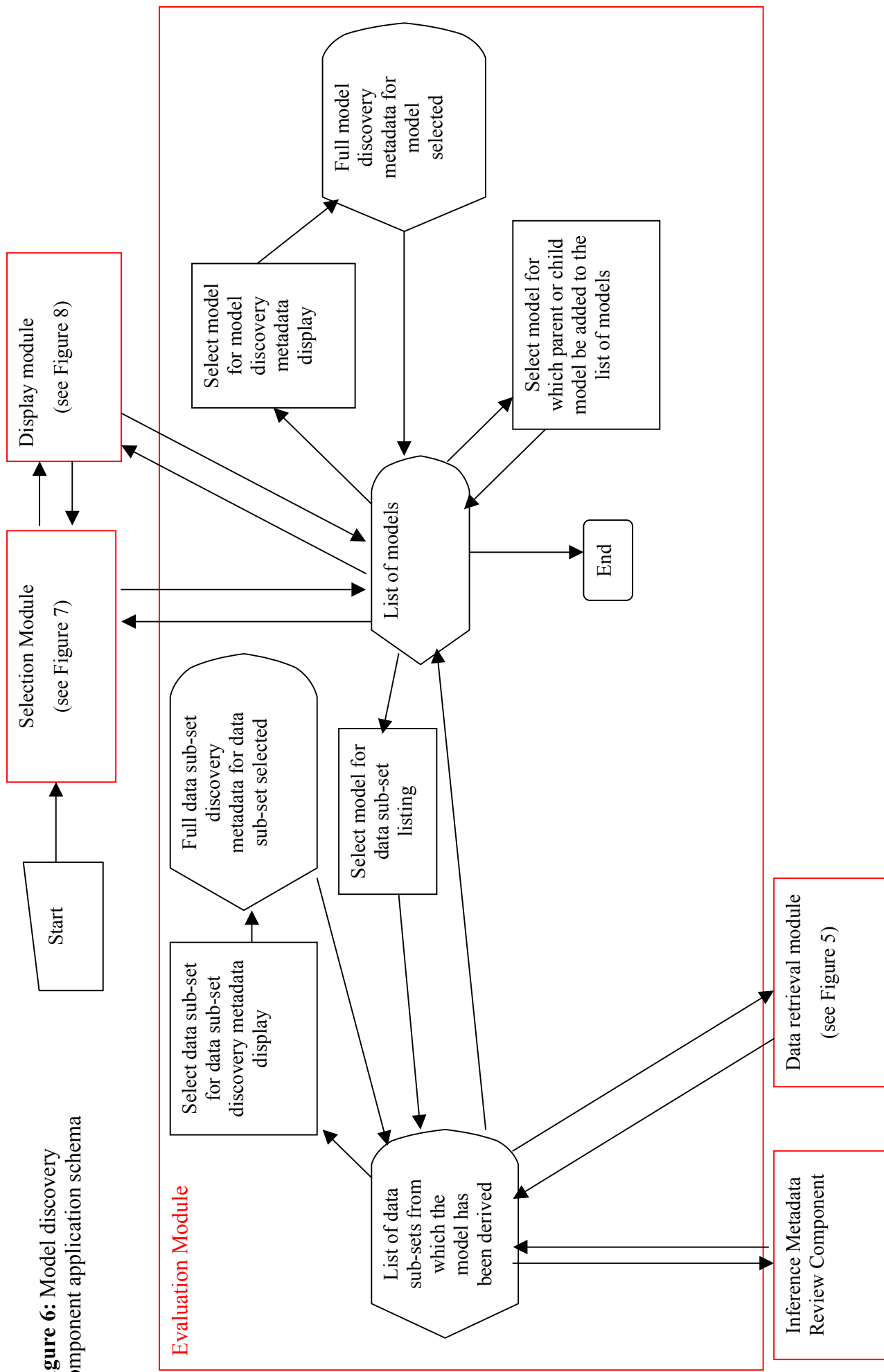


Figure 7: Model selection module schema

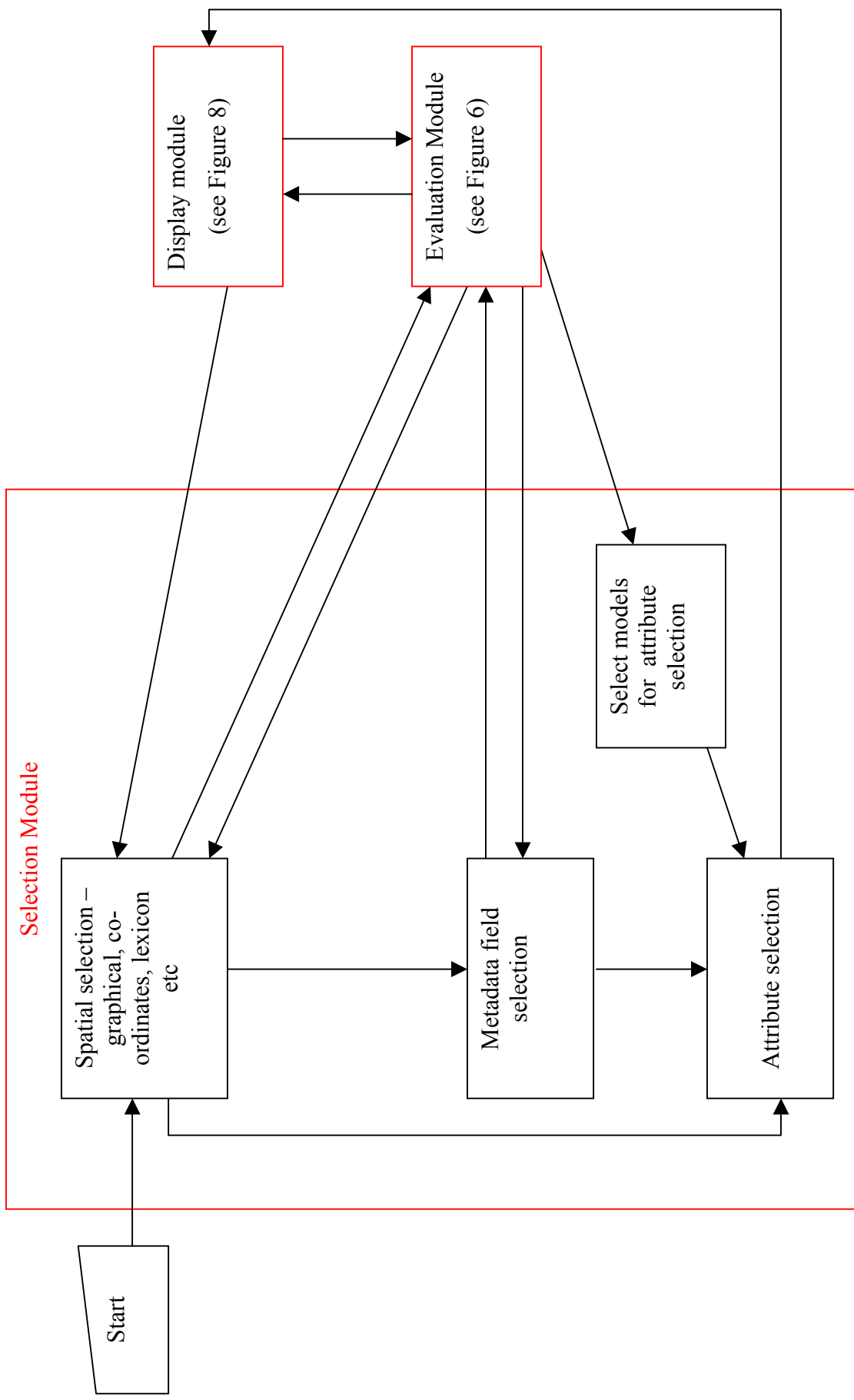


Figure 8: Model display and retrieval module schema

