Digital literacy training

**NVivo12 Pro Advanced**

Advanced Coding and Data Analysis

2020
# Table of Contents

**NVivo Advanced Coding and Data Analysis** .................................................. 1

**Deciding on a Coding Scheme: Word Frequency Queries** ............................. 1

  - Word Frequency Results ............................................................................. 2
  - Excluding Irrelevant Words from Searches ............................................... 2
  - Word Clouds ................................................................................................. 3
  - Tree Map ..................................................................................................... 3
  - Cluster Analysis .......................................................................................... 3
  - Looking at the Results in Context ............................................................... 4

**Initial Mass Coding: Using Text Search Queries** ....................................... 4

  - Text Query Results Preview .................................................................... 5
  - Saving the Query Preview Results as a Node ........................................... 5
  - Expanding the Coding Context of References in Results or a Node ........... 5
  - Automatically Expanding the Coding of References in a Node .................. 6
  - Word Trees ................................................................................................. 6

**Analysis** ....................................................................................................... 7

  - Coding Queries ......................................................................................... 7
  - Basic Coding Queries using the Query Wizard ......................................... 7
  - Advanced Coding Queries ....................................................................... 8
  - Editing a Query ......................................................................................... 8
  - Compound Queries .................................................................................. 9
  - Creating a Compound Query ................................................................... 9
  - Matrix Coding Queries ........................................................................... 10
  - Creating a Basic Matrix Coding Query ..................................................... 10
  - Editing Node Matrices ............................................................................ 11
  - Saving Matrices ....................................................................................... 11
  - Exporting Node Matrices ....................................................................... 11
  - Crosstab .................................................................................................... 12
  - Cluster Analysis ....................................................................................... 12
  - Editing Cluster Analysis Results .............................................................. 14

**Maps** ........................................................................................................... 14

  - Concept Maps .......................................................................................... 14
  - Project Maps ............................................................................................. 15

**Graphs** ......................................................................................................... 16

  - Explore Diagrams ..................................................................................... 16
  - Comparison Diagrams ............................................................................. 16

**Exporting Results and Project Items** ......................................................... 17

  - Exporting the Codebook .......................................................................... 17
  - Exporting Classification Sheets ................................................................. 18
  - Exporting Nodes ...................................................................................... 18
  - Exporting the List View of the Nodes Area .............................................. 18

**Collaboration** ............................................................................................... 18

**Interrater Reliability** ................................................................................... 19

**Other resources** .......................................................................................... 20

---

Digital literacy training
NVivo Advanced Coding and Data Analysis

This course will show you some more advanced NVivo features and how to use the most commonly needed NVivo tools to analyse your qualitative data. For further information about NVivo’s full capabilities, please refer to the NVivo 11 help website (NVivo 12 comprehensive is not yet available): help-nv11.qsrinternational.com

This session will show you how to:

- use Word Frequency Queries to decide on a coding scheme
- initially mass code your text data using Text Search Queries
- analyse and explore your data further using Queries: Coding, Compound and Matrix
- summarise your data using Crosstab
- analyse your data using Cluster Analysis
- Export results and project items
- create and edit special NVivo Graphs
- create Maps to visualise your project structure
- collaborate in an NVivo Project

Open the Sample Project to explore how NVivo can be used to analyse your data. When you open NVivo, at the top of the Welcome Screen there is an option to open a new Sample Project. Select the sample project which is a two-year study (2008-2009) documenting community perceptions of development and land-use change on coastal communities in the Down East area of Carteret County, North Carolina, USA.

Important: Save the Project under a new name so you can alter the data without changing the sample project for other computer users. Go to File choose Copy Project and save the file as Sample Project (2) in your My Documents folder. Saved Projects can be opened on any computer with NVivo and can be altered by any collaborator (a record of who changes what is kept according to User ID).

Deciding on a Coding Scheme:
Word Frequency Queries

In the initial stages when you are still deciding on a coding scheme, a Word Frequency Query could be useful. This will produce a list of all the words that occur most often in the text documents you have (including memos, transcripts etc. unless you specify otherwise). You can search the entire content of your qualitative data, or you can narrow it down to specific groups of people or individual questions, etc.

Example: Word Frequency Query for Interview Question 4: “Community & Environmental Change”.

On the Explore tab, click on Word Frequency. This will open a new Detail View with the Word Frequency options. Choose where to search in the Search in area. For this example, click Selected Items and then click the + next to the Nodes folder to see the other folders within it. Then click on the Autocoded Responses folder and then the Auto coded Interview Questions folder and tick the box for Question 4. Then decide if you want to limit the number of results (by default to the 1000 most frequent) and/or change the largest letter-length of the words you want in the results (default is 3 letters so you don’t get words like “to”, “a”, “an” or “I”). In the Grouping Area, decide on the grouping criteria by sliding the indicator (e.g., With stemmed words - see table below for descriptions of the 5 levels).

Note: The higher the similarity, the longer it will take to run.

Digital literacy training
**The five Text Match levels:**

<table>
<thead>
<tr>
<th>Grouping Level</th>
<th>Returns</th>
<th>Example (sport)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact matches</td>
<td>Exact matches only</td>
<td>sport</td>
</tr>
<tr>
<td>With stemmed words</td>
<td>+ Words with same stem</td>
<td>sport, sporting</td>
</tr>
<tr>
<td>With synonyms</td>
<td>+ Synonyms</td>
<td>sport, sporting, play, fun</td>
</tr>
<tr>
<td>With specializations</td>
<td>+ Specializations</td>
<td>sport, sporting, play, fun, running, basketball</td>
</tr>
<tr>
<td>With generalizations</td>
<td>+ Generalizations</td>
<td>sport, sporting, play, fun, running, basketball, recreation, business</td>
</tr>
</tbody>
</table>

If you want to save the Query details and run it again later from the **Queries** folder in the **Search Area**, click **Add to Project** and give the Query a name then select **OK**. Then click **Run Query**.

**Note:** A Text/Word Search Query does not search text within PDFs created by scanning paper documents (each page is a single image). Consider using optical character recognition (OCR) to convert the scanned images to text before importing.

**Note:** Only words in the transcript fields are included in Queries when searching audio and video Data Files and only in "codable" fields in datasets.

---

**Word Frequency Results**

There are several tabs on the right-hand side of the **Detail View** which present the results in different ways. From these you can decide on an appropriate coding scheme and even begin to create Nodes. The **Summary** tab shows you a list of all the most frequent words, their **frequency Count**, other summary data and if you have chosen a **Grouping** level higher than **Exact matches**, you will also see the other words it has included as belonging to each Word Count (e.g., the word count for Change also includes that for Changed, Changes, Changing). There are also tabs for graphical representations of your Query: **Word Cloud**, **Tree Map** and **Cluster Analysis** (the latter only appears if there is enough data to run it).

---

**Excluding Irrelevant Words from Searches**

In the results you might see words that you do not want to include in the Query search in future, such as "any". Right-Click on the word in the results and select **Add to Stop Words List**. If you have included more than exact matches in the text match level section (i.e., >level 1), you must remove any words in the box which you still wish to be included in future searches. When you are ready, click **OK**. Every time you run a Query from now on, these words will not be included in results. You can edit these in **Project Properties** at any time: on the **File** tab under **Info** select **Project Properties** and on the **General** tab click the **Stop Word List** button and edit the words to exclude.

**Note:** it is a good idea to check the **Stop Words List** before you run any **Text Searches** or **Word Frequency** Queries. There may be some words on the list that you **don't want NVivo to ignore** in these Searches in Queries. You will have to remove these in the **Project Properties** first, because asking NVivo to search for these words specifically does **not** override the Stop Words List.
Word Clouds
The **Word Cloud** tab displays up to 100 of the most frequent words in your text data, varying in font size: the largest are the *most* frequently occurring, and the smallest the *least* frequent.

Tree Map
The **Hierarchy Map** tab displays up to 100 of the most frequent words in rectangles of varying sizes: the largest rectangles are for the most frequently occurring words, and the smallest for the least.

Cluster Analysis
The **Cluster Analysis** tab displays up to 100 of the most frequent words in a tree-diagram describing the co-occurrence between these words. The closer the words are on the branches of the tree map, the more frequently they co-occur. This can help by providing more context.
Looking at the Results in Context

If you think that one of the words in the results might be a good Node to have in your Coding Scheme, you can **view each word in the results in context**. Double-click on the word in the Summary table/Word Cloud/Tree Map, or right-click and Choose **Run Text Search Query**. This runs a Text Search for the word (for exact matches) or words (for grouped matches) and opens a new tab in the **Detail View** containing all the relevant references containing the word(s).

**Note:** You can also create *initial* Nodes from Text Searches like these (as described in the section below). Of course, they will have to be "cleaned up" afterwards.

Initial Mass Coding:
Using Text Search Queries

**Text Search Queries** can be used to help start the coding process for a large amount of textual data by finding instances of text matching specific **search terms** based on your pre-defined coding scheme (**internal** text data files ONLY). Information coded to Nodes this way can always be modified later through uncoding if necessary.

**Note:** Consider checking the spelling before running Queries. You will find the **Spelling** button on the **Home** tab.

**Example: Create an initial, automatically coded Node for Water Quality.**

On the **Explore** tab, click on **Text Search** and this will open up a new **Detail View** window with the Query options at the top. Choose where to **Search in** (e.g., Selected Folders) and tick the boxes for the folders you want to search in (e.g., in Interviews and Surveys), then click OK. In the **Search for** area, enter the word(s)/phrase you want to look for. You can create more complex Queries using the special features displayed when you click the **Special** button:

**Wildcards:** If you want to search for words that have alternative spellings (e.g., colour/color), or are commonly confused with similar words (compliment/complement), Wildcards can help to find every possibility. The Wildcard * is used to stand in for one or more missing characters (e.g., g*t will find get, great, gradient etc.). The Wildcard ? is used to stand in for a **single** missing character (e.g., g?t will find get, gut, got etc.).

**Fuzzy Search:** Also useful to pick up misspellings and find words with similar stems. The symbol ~ behind a word will look for similar words (e.g., analyze~ will find analyzes, analyse, etc.).

**Note:** If you use the above functions you will be limited to **Exact match** only.

- **Boolean Operators:** When you want to find a specific concept you can use a combination of words and/or excluding a word(s) in your search.
  - **AND** (or &) will search for BOTH terms (water AND quality).
  - **OR** (or |) will search for EITHER of the terms (water OR lake OR sea OR river).
  - **NOT** (- or !) will search for the first term and EXCLUDE any results that also contain the second (water NOT tap, water -tap).
  - **REQUIRED** (+), will search for items containing the required term and the second term is optionally included in results if and only if it occurs together with the first term (+Water river).
  - **Double Quotes** " " should be used to enclose a group of words to search for a specific phrase. If you search for a phrase, the results will not include words with the same stem. For example, if you search for "alternative energy", you will not be able find "alternative renewable energies".
  - **Round Brackets** ( ) can be used to group clauses to form sub Queries.
• **Near**: If you want to find words that co-occur but you want to ensure they are mentioned in the same sentence rather than several sentences away, you can specify the distance either side of the required word (e.g., “water AND quality”~20) will find instances of water and quality within 20 words of each other only.

**Note**: you need to enclose the search terms in **double quotes** before adding the ~ for a Near Search. In addition, the tilde (~) denotes proximity ONLY in this search.

For this example, enter “water AND quality”~20. **Spread Coding** to a **Broad Context**, so you can see enough information when you review the success of initial coding. Decide on the **matching criteria** (e.g., With stemmed words). Click **Add to Project** if you want to save the Query parameters and make it easier to rerun it later if you obtain new data. Then click on the drop-down menu for **Run Query** (you can always **Save Results** as a new Node or merge the results into an existing one, later if the results are what you wanted).

### Text Query Results Preview

In the preview of the Query results, there will be tabs on the right of the **Detail View** for:

- **Summary**: A list of all the Data Files that contain the word/phrase results.
- **Reference**: The results are opened as a Node preview and the word or phrase is shown with a narrow, broad or custom context (which you can expand: see previous workshop notes).
- **Text, Picture, Audio, Video, or Dataset**: Displays the results found in each type of Data File (only relevant tabs are available).

### Saving the Query Preview Results as a Node

In general, if you run any Query and decide retrospectively that the results in the preview would be valuable to save as initial coding for a Node, click on the Query results in the **Detail View**, then on the **Node Tools** tab click the **Store Query Results click rolled** and select Store Query Results. From the **Option** list, choose to either:

- Create the results as a new **Node or Case**
- Merge the results into an **Existing Node or Case**

Another method is to right-click somewhere inside the Query results in the Detail View and select **Store Query Results**. Give the new node a **Name** and optional **Description**, then click OK.

**Note**: by default new Nodes are created in the **Results** folder unless you choose another location (e.g., **Nodes**).

**Note**: NVivo will just code the word(s) found in that Query unless you expand the coding.

### Expanding the Coding Context of References in Results or a Node

Sometimes when you have run a Query, the context for the references in the Nodes can be quite narrow (e.g., only a few words) and you need to see more text to assess whether more should be coded to the Node.
Within the **Query Results** or **Node** you can right-click on the chosen reference in the **Detail View**, and select **Coding Context** and choose the level you want:

- **Narrow** to see 5 words, 5 seconds or 5 percent on either side of the coding reference, or (for datasets) see all the text in the cell.
- **Broad** to see the surrounding paragraph, cell or row, or see 20 seconds or 20 percent on either side of the coding reference.
- **Custom** to define specific reach settings
- **Entire File**

Then click **OK**. More context will appear in **grey font** on either side of the coded text which is differentiated by **black font**. If you want to expand the coding context for the entire Node, select all the text (either using the mouse or Ctrl+A), then right-click on the selection and follow the instructions above.

**Automatically Expanding the Coding of References in a Node**

If you would like to expand the coding itself to include a broader/narrower context, when you are in the **Node** you can right-click on the chosen reference in the **Detail View** and select **Spread Coding** instead and choose the level you want. Again, you can do this for the entire Node as explained above.

**Exercise: Run a Text Search Query for synonyms of the words Pollution OR Contamination. Do not Spread Coding and do not save the results in order to create a Word tree.**

### Word Trees

**Word Tree** tabs are produced when you run **Text Search Queries** (only if you choose NOT to spread coding OR save the results). The branches of the **Word Tree** represent the various contexts in which the word or phrase occurs. It can be useful for finding recurring themes of phrases that surround the word.

You can click one branch to highlight the other end of the branch to see the actual context. When on the **Word Tree** tab, a main **Word Tree** tab appears at the top of the window. This allows you to change the **Root Term** (e.g. to polluted), the **Branch Order** (e.g., by **Number of Matches**) and **Context** level (the number of words surrounding the root term).

[Diagram: Pollution or Contamination - Results Preview]

You can also open the **Data File** where a specific quote in the Word Tree occurs, by right-clicking on a branch, then selecting **Run Text Search Query**.
Analysis

When you have completed your coding, there are 4 types of Queries to help you explore your data in more depth and also to quantify it:

- **Coding**: explore your data more in depth based on how you coded it (e.g., What did people say about *Water Quality* and *Environmental Impacts*? What did they say about *Impacts* independent of *Water Quality*?).

- **Compound**: combine two Text and/or Coding Queries to explore your data further (e.g., Did people mention terms related to "health" in the Node *Water Quality*?).

- **Matrix**: creates a cross-tabulation of your Nodes/Attributes to look for patterns and quantify your results (e.g., how many people said positive things about *Water Quality* and how many said negative things? What themes were more prominent in response to each survey question?).

- **Crosstab**: like a Matrix Query but can only create a cross-tabulation of Nodes by Attributes, and you cannot select specific attribute values.

Note: before running any Query check to make sure that Aggregation for the Parent Nodes in your analysis is turned ON if you want to include results for their Child Nodes, and OFF if you do not want to include them. You must do this for any grandchild and great-grandchild Nodes if you want those included as well. See notes from previous session regarding Aggregation.

Coding Queries

Basic Coding Queries using the Query Wizard

You can run a Coding Query to search for content based on how it’s coded at specific Nodes. You can find content coded at:

- **A particular Node** (e.g., *Water*) [same as opening a Node to view its contents]
- **All of these Nodes** you specify (e.g., *Water Quality* AND *Fishing* OR *Aquaculture*)
- **Any of these Nodes** you specify (e.g., *Habitat* OR *Forest* OR *Landscape*)
- **Any Node where** there are particular Attribute values (e.g., all the responses of the Person classification attribute of *Gender* equals *value Female*).

Example: How many and which references are coded at both *Water Quality* AND the specific Economic concern of *Fishing* or *Aquaculture*?

On the Explore tab, click Query Wizard, then click the Search for content based on how it is coded, and then click Next. Choose All of these Nodes and tick Water under the *Natural Environment* Node, and *Fishing* or *Aquaculture* under the *Economy* Node.

Note: if you tick the Automatically select descendent Nodes (meaning child Nodes) it will look for co-occurrences of all the child Nodes under *Fishing* or *Aquaculture* as well, and as most of these are mutually exclusive, will return no results. This parent Node has Aggregate coding from child Nodes ticked, so it will include information contained within these child Nodes anyway (see in the Node’s Node Properties).

Then click Next and choose to look for coded content in Items in Selected Folders to restrict the Query, click the Select button and choose Interviews and Surveys folders. Click Next, then choose whether you want to Run this Query Once or Add this Query to Project (and run it). If you choose to add it to your project, you must enter a name and (optional) description. Please note it will be saved in the Queries area. Then click Run. All the results that satisfy the Query conditions will be displayed in the Detail View.
Advanced Coding Queries

If you want to use more advanced Coding Queries that have features that are not available via the Wizard (e.g., find coded content NEAR content coded at other Nodes), you can use the Coding function on the Explore tab.

Example: What do respondents 40 years of age and over say about Community Change?

On the Explore tab, click Coding and then a new Detail View will open with Query options. Search in the Selected folders Interviews and Surveys. From the drop-down menu select All of the following are true (for the AND Boolean function), then Coded At and Any case where from the next drop-down menus (to specify Attributes). Click on the "..." box to select the Attribute conditions for the Query. Expand the list for Person and select Age Group then click OK. Now you have a new drop-down menu on the right from which to select the logical argument: choose (>) Greater than and from the next box select the value 30-39.

Note: NVivo doesn’t recognise age groups in this format as being numerical data so it doesn’t know what order they should be in, it just assumes you have them in the right order.

Add the next condition to the Query by clicking the + button. Choose Coded at and All Selected Codes or Cases then click the "..." box to choose the Node Community Change. Spread Coding to the Broad context and click Run Query.

You can also add conditions such as NEAR and Coded by any by clicking on the drop-down menu next to the +/- buttons to the right of a row before adding the next coding criteria.

The results show you everything that people 40 years and over said about Community Change.

Editing a Query

If you make a mistake in a Query, you can always amend it in the Detail View, however if you have already closed this you can click the Last Query icon on the Explore tab to go back to the Query criteria.

You can edit the search criteria in the Search for content matching these criteria box:

- Change the order of processing, by selecting an item from the criteria list and click the required up or down arrow next to the +/- buttons.
- Remove an item in the criteria list, by selecting the item and then click the “-” button.
- Edit an item in the list, by editing the criteria in the Detail View.

Note: If you edit a Query that has been added to the project, this will overwrite the original Query. If you want to duplicate a Query and make changes to it (e.g., the same Query but with different Attribute Values), you can always copy and paste the Query in the Queries folder of the Search Area. This will create a duplicate Query with a numbered extension in brackets that you can rename and change at any time.

Exercise: Find out what the interviewees and survey respondents say about Real Estate Development AND Water Quality. Use a NEAR search to allow results that almost overlap in the Nodes.
Compound Queries

Compound Queries can be used to refine a text search by:

- Combining 2 Text Search Queries to find where one term precedes another (e.g., with related to bad before Quality as opposed to good before Quality).
- Combining 2 Coding Queries to search for content coded at one Node, near content coded at another just in case relevant content doesn’t overlap (e.g., where do climate change and sea levels Nodes co-occur within 20 words).
- Combining a Text Search and Coding Query to search for text surrounded by or near a Node (e.g., do people mention words similar to sceptical in/near content coded at the Node climate change).

Creating a Compound Query

Example: Search for mentions of “weather” near the Node Natural Environment.

On the Explore tab, click Compound. This opens a new menu window only, not a Detail View like the other searches. In the Compound Query window select the first Subquery 1 type: Text Search, then click the Criteria button and enter the Search for criteria (e.g., weather). Choose a match level (e.g., the highest: Similar) and select to Search In: Text, Annotations or Both (e.g., Text). Then click OK. In the next box, choose the appropriate search terms (e.g., Near Content):

- AND / OR / NOT (see Advanced Text Search Queries)
- NEAR content - coded at Node A near other content coded at Node B within # words (before or after).
- PRECEDING content - coded at Node A when it precedes content coded at Node B.
- SURROUNDING content - coded at Node A where it is surrounded by content coded at Node B.

If you choose NEAR or PRECEDING Content, click on the Options button to specify the proximity level required:

- Overlapping - when coding at both Nodes/context overlaps.
- In Custom Context - within a specified number of words, surrounding paragraph/cell/row, minutes or percentage.
- In Same Scope Item - within the same item, such as the same document or dataset.
- In Same Coding Reference - within the same coding reference (only available when Nodes are included in the Query scope).

For this example, choose In custom context then click Specify button, and select 20 for the Number of words. You can also choose how much content you want to gather in the resulting Node/preview (e.g., tick all 3 boxes) in the Options. You can tick:

- Finds for first search item - gathers content matching the criteria defined for the first search item only.
- Finds for second search item - as above for the second search item only.
- Content between finds - displays all content between the first and second search items. Note: This option is only available if both of the above check boxes are selected and Overlapping is not the chosen Proximity option.
- Compare proximity between text and non-text items - gathers content from different components of an audio/video or picture File (e.g., Query both the video and the video transcript. If you do not select this check box, the Query will evaluate proximity within the same type of content: either text or media).
In the SubQuery 2 area, select the second Query type from the drop-down menu (e.g., Coding) and click the second Criteria button and enter the Search for content coded at criteria: All of these Nodes (e.g., Natural Environment). You have the Option to set the Scope of your Query from the Search In list (e.g., Items in selected folders), select the items you want to include by clicking the Select button (e.g., Interviews and Surveys). Click OK. Click the Query Options tab to define the spread coding options and preferences for storing the results (e.g. broad). Click Run.

Matrix Coding Queries

You can use Matrix Coding Queries to cross tabulate the frequencies of Nodes and/or Attributes in order to:

- compare what different demographic groups have said about an issue (e.g., are different Townships concerned about different issues?).
- explore the amount of overlap between different themes (e.g., is there a potential relationship between Community Change and the Economy, Infrastructure, Real Estate Development, and Policy?).
- compare attitudes (e.g., do Males have more Positive and less Negative attitudes than Females within the Environmental Change Node?).

They are also particularly useful for further descriptive and inferential statistical analysis using chi square for example (when cells are independent). Most often Nodes and/or Attributes are used in the rows and columns of the matrix, but you can also select Data Files. In addition, you can create matrices that only relate to content coded within a specific Node (e.g., a specific survey question).

Creating a Basic Matrix Coding Query

Example: Do Commercial Fishers have more Positive, Negative, Mixed or Neutral attitudes than Recreational within the context of Water Quality?

On the Explore tab, click Matrix Coding. This will open another menu window with a Detail View. In the Rows box, add Project Items by clicking the + for the box and choosing either Select Items (e.g., Nodes or Cases) or Select Attribute Values (e.g., Age Groups).

To Select Items, just tick the boxes of the items you want the rows to represent and click OK. Alternatively, you can click and drag Nodes into this box from the Nodes list view (e.g., With the Nodes list view visible, select all the Attribute child Nodes: Mixed, Negative, Neutral and Positive then click and drag them to the Rows box and let go). Add columns in the same way by clicking on the + for the Columns tab.

To add Attribute Values you must click on the + but counterintuitively choose Select Items. This allows you to choose individual attribute values by ticking boxes rather than going through multiple steps and defining logical arguments (e.g. Recreational Fishing = Yes). Click on the Case Classifications folder name (do not tick the box), then in the Person Classification find and tick both of the Recreational Fishing and Commercial Fishing Yes value boxes.

Choose where you want to search (e.g., in this example search in Selected Items and choose the Water Quality Node from the Nodes folder under Natural Environment Nodes) then click OK. Choose if you want to Add to Project then click Run Query.

Note: By default, the cells of the table will represent the conjunction between rows and columns (the AND Boolean operator). You can change this if you need to by clicking on the Coding at Rows drop-down menu and choosing from OR, NOT, Near, Preceding and Surrounding.

Note: If you want to display the results as row or column percentages, the results it displays will refer to the number of words coded in that cell as a percentage of total words coded in that row or column.
Within the results of a Matrix Query you will see a cross-tabulation known as a Node Matrix. You can double-click on a cell to open the summary of all the references relating to that cell (like running a Coding Query for that cell). In this sense, Matrix Coding Queries can be useful alternatives to Coding Queries. Matrices can be viewed as charts, by clicking on the Charts tab in the Detail View. You can change the chart type, title, gridlines and rotation (see more on Creating and Editing Charts on the NVivo help website).

Editing Node Matrices

You can click and drag columns wider or narrower to see more content. Columns and rows can be sorted in ascending or descending order, by selecting the column/row of interest, on the Matrix Tools tab click on Sort & Filter then click Sort by Column or Row. Each column and row can be filtered this way also or alternatively by clicking on the appropriate filter (funnel) symbol for the relevant row or column. NVivo will show or hide columns/rows based on the filter (e.g., hide rows where number of responses = 0). If you need to transpose columns and rows, click on the Node matrix then in the Matrix Tools tab, click Transpose.

You can also apply shading to cells to make it easier to see patterns. On the Matrix Tools tab select one of the cell shading options displayed (e.g., blue and white). The darker/hotter the shading colour the more coding in that cell. You can also change the information displayed in the cells here (e.g., Number of Nodes Coded as Person).

Note: By default, the Node Matrix displays the number of References in each cell (i.e., the number of text sections coded), and multiple references may be from the same Case OR multiple collaborators coded the same content. To see the number of Cases for each cell (often the most useful metric), on the Matrix Tools tab click the Cases Coded drop-down menu and choose Person.

Note: You can also display the row and column percentages, however please be aware that this will only ever relate to the number of words coded not the number of Cases or references.

Saving Matrices

You can save the matrix retrospectively in the Node Matrices folder or the Query Results folder within the Search area for future reference. On the Matrix Tools tab click on the Store Query Results icon, and select the location you want to store it in. Give the matrix a Name and (optional) Description, then click OK.

Exercise: Which Townships are more concerned about Habitat, Landscape and Water Quality Nodes. Include all the townships in rows, and the 3 Nodes in columns.

Exporting Node Matrices

You can export a Node Matrix to an Excel, text, or SPSS file so you can quantitatively analyse it further or create graphs in Excel. Right-click on the Matrix you want to export and select Export Node Matrix. Select a location to save the file and change the name if necessary, and in the Save as Type box choose the appropriate file format. Click Save.

Handy Hint: Excel is probably a better program in which to create charts from your quantitative NVivo data, as you will have far more control over their look and content.
Crosstab

There is a new feature in NVivo 12 called **Crosstab** which is designed to be a quick way to check the spread of coding across cases and demographic variables. It essentially does the same thing the Matrix Coding Query with some limitations: you can only compare **Nodes** and **Attributes** and you are not able to select specific Attribute values.

**Example: Which age groups have more Positive, Negative, Mixed or Neutral attitudes within the context of Water Quality?**

On the **Explore** tab, click **Crosstab**. This will open another menu window with a **Detail View**. In the **Codes** box, add Nodes (**Note**: you cannot add Cases) by clicking the + for the box and tick the boxes for the Nodes you want to include in the table (e.g., **Mixed**, **Negative**, **Neutral**, and **Positive**).

To add **Attribute Values** select the appropriate Classification (e.g., **Person**) and then choose an Attribute from the **Attribute 1** drop-down menu (e.g., **Age Group**). You can also select a 2nd Attribute in the **Attribute 2** drop-down menu if you wish. **Note**: you cannot select individual Attribute Values, all will be displayed in the table. You can however choose to include **Unassigned** and **Not Applicable** values by ticking the relevant box in the menu.

Select where you want to search (e.g., in this example search in **Selected Items** and choose the Water Quality Node from the **Nodes** folder under Natural Environment Nodes) then click **OK**. Choose if you want to **Add to Project** then click **Run Query**.

You can edit a Crosstab Matrix in a similar way to a Matrix Query result.

Cluster Analysis

Cluster Analysis is helpful for identifying patterns in your data by grouping Files or Nodes into clusters (maximum of 10 by default). This can help you to visualise the similarities and differences between selected **Files** or **Codes**1. Similarity metrics can be based on the words used, coding conducted or attributes applied. It uses the **farthest neighbour** (a.k.a. complete linkage) hierarchical clustering method.

You can cluster the following project items:

- **Files, Externals** and **Memos** (e.g., to identify similarities between journal articles or Cases of each file represents an individual Case).
- **Codes** (e.g., to identify similarities between your Nodes or between Cases).

You can cluster selected Files or Codes based on:

- **Word similarity**: based on the number of words they have in common. **NOTE**: stop words are excluded when using this measure.
- **Coding similarity**: Data Files are coded based on the number of Codes they have in common, whereas Codes are coded based on the number Files they have in common. **NOTE**: this measure includes ALL of your internal and external Data Files which contain coding, and you cannot select a subset.
- **Attribute value similarity**: Data Files are coded based on the number of File Classification Attribute values they have in common. Only **Cases** can be coded based on their Case Classification Attribute values. **NOTE**: it is best to select a File or Case Classification to cluster for this measure to ensure they all have the same classification attribute values with which to compare them.

---

1 Here **Codes** refers to both Nodes and Cases
### Table of kinds of questions different clustering methods can help you answer:

<table>
<thead>
<tr>
<th>Cluster What?</th>
<th>Data Files</th>
<th>Nodes/Case Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word Similarity</strong></td>
<td>Which articles are similar in content? Which journals tend to publish similar content?</td>
<td>How similar are selected Nodes in terms of the content coded? Are they related? Are they so similar they should be merged? Which people (Cases) hold similar views according to their survey responses?</td>
</tr>
<tr>
<td><strong>Coding Similarity</strong></td>
<td>Which articles have similar themes? Are there groups of people who are concerned about similar topics?</td>
<td>What group of themes tend to be mentioned together in a lot of articles/interviews? Which authors (if Cases) tend to publish together?</td>
</tr>
<tr>
<td><strong>Attribute Value Similarity</strong></td>
<td>Which articles have similar reference classifications (author, journal, etc.)? Are there distinct groups to analyse separately?</td>
<td>Which Cases are similar in terms of their demographics? Are there distinctive groups of people you might wish to analyse separately?</td>
</tr>
</tbody>
</table>

You can choose from 3 similarity measurements:

- **Pearson’s correlation coefficient**: Ranges from -1 = least to +1 = most similar.
- **Jaccard’s coefficient**: Ranges from 0 = least similar to 1 = most similar.
- **Sørensen’s coefficient**: as above.

**Example: Conduct a Cluster Analysis to see which Nodes are similar in terms of the words coded.**

To run a Cluster Analysis, on the **Explore** tab, click the **Cluster Analysis** icon to open the **Wizard**. Select to cluster by **Codes** (to cluster by your thematic **Nodes**), then click **Next**. Then in the Nodes box, click the **Select** button and choose the **Nodes** you wish to cluster (e.g., all the parent, child and free Nodes in Nodes folder except for **Memorable Quotes** and the parent **Attitude** Node). Then select how you want to cluster them from the drop-down menu (e.g., by **Word Similarity**), and then choose the **Similarity Metric** (e.g., Jaccard’s). Click **Finish**.

![Nodes clustered by word similarity](image-url)
The **Diagram** tab shows you a dendrogram (by default) which indicates which Nodes are most similar in terms of the words in their coding. Those that cluster together on a branch are more similar than those further apart. In the dendrogram above, it is interesting to note that words coded in the *Local Connection* and *Local Identity* are also similar to words used in a *Positive* context.

On the **Summary** tab you can see a metric of how similar each pair of Nodes are, from highest to lowest. Items with an index value close to 1 are very similar and as such will appear closer together on the diagram. If you want to see what words Nodes in a cluster have in common, click on the Node names to highlight them, right-click on the selection and choose **Run Word Frequency Query**.

If you want to see the Attribute values of a File or Case in a cluster analysis (e.g., to see if there are demographic similarities between items that cluster together) select the File or Case, right-click on it and select **Item Properties**.

**Editing Cluster Analysis Results**

You can change the type of cluster graph when you are in the results area, in the **Cluster Analysis** tab that appears in the top menu list. Click on the appropriate icon to change the type.

**Maps**

There are 3 types of Maps in NVivo12:

- **Mind Maps** – For organising your Coding Scheme (and creating Nodes). See previous course notes.
- **Concept Maps** – A more free-form way of visually representing the relationships between Project Items.
- **Project Maps** – To get NVivo to visually represent specific Project Items and how they relate to each other within the Project itself.

Dynamic Models can be changed, and Static Models are read-only (all links are removed, but it’s useful for keeping snapshot records of your progress). All newly created models are dynamic.

**Concept Maps**

Concept Maps are free-form rather than strictly hierarchical and allow you to create diagrams with a variety of shapes and connectors, and elements that are connected to Project Items (i.e., you can open the project items directly from the map). Concept Maps are saved in the Maps area of NVivo.

On the **Explore** tab click **Concept Map**. Enter a name for the model in the **Name** box, and an (optional) **Description**, then click **OK**. In the **Detail View** click and drag the shapes you want to add from the window pane on the left. Consider using different shapes to denote different elements of your project (e.g., circular for central and square for peripheral themes). Double-click on the shapes to enter labels. You can also change the colour of individual shapes by clicking on the one you want to change then select the colour and/or border colours you want on the appropriate drop-down menus on from **Concept Map** tab.

To connect shapes to each other, click the **Connector** button on the **Concept Map** tab which changes the cursor to a connector. Click and drag from one shape to another to connect them. If you want to change the directionality of the connector, click on the connector and then the **Change Connector** button on the tab and select the direction you want. If you want to label the connection, simply double-click on the connector and a text box will appear. To get your normal cursor back, just click **Pointer**.
If you wish to add a shape which is a link to a Project Item, click the Add Project Item button to bring up a menu and tick all the items you want to add, and click OK. You can also click-and-drag project items from the List View to the map.

You can delete a shape by selecting the one you want to delete then on the keyboard click Delete.

Add a note to a model

You can add notes to models which allow you to enter more text: on the Model tab, click the Note shape from the Shapes drop-down menu. Double-click the note and in the Text box, enter your notes. Click OK.

Project Maps

Project Maps can be used to get NVivo to show you a graph of the final coding structure of your Project, or how Project Items are connected to each other (e.g., you can look at all the Nodes associated with one Case or File). You can also click on one of the items and asked to see a list of all associated project items (e.g., Child Nodes, Cases, Files, Links), to decide if you want to include them in the map. Concept Maps are saved in the Maps area of NVivo.

On the Explore tab click Project Map. Enter a name for the model in the Name box, and an (optional) Description, then click OK. On the Project Map tab, click on the Add Project Items icon to select the items you wish to include in the map (e.g., the Economy Node and all its child Nodes). Then click OK. On the Project Map tab you can deselect any connectors you wish to hide from the map in the Connectors Area of the Project Map tab (e.g., Links).

A quick way to see the Project Items that are associated with an Item already in the Map (and to show this association on the map also), click on the Item and on the Project Map tab, and then click Show Associated Items icon at the top of the window. A new window appears on the left-hand side of the Detail View which lists all valid items you can include. You can click and drag the items you want to add from that pane.
Graphs [OPTIONAL]

Explore Diagrams

You can also create special graphs to show how project items are connected to one specific File, Node or Case. This is very similar to a Project Map, only the latter can show you how the other Project Items relate to each other as well. For example, a Project Map will show you how one Node is connected to other Nodes, whereas an Explore Diagram will only show you what project items other than Nodes are related to Nodes (e.g., Cases, Files etc.).

**Example: Create an Explore Diagram of the Project Items related to Balance.**

Select the File, Node or Case you want to create a graph for in the List View of the Codes, Cases or Data area (e.g., select Balance). On the Explore tab, click on the Explore Diagram icon to generate the graph. You will see a new Graph appear in the Detail View, which shows you all the project items in a related to it. As this is a dynamic graph, you can click on any one of the icons to open-up that Project Item. You have extra options such as showing or hiding information on the Explore Diagram Tools tab (e.g., hide Files coded or Memo Links). You can also change the focus to another Node within the graph itself by right-clicking on a Project Item object, select Change Focus and a graph will be generated with that Node at its centre. If you want to go back to the original, click the Back button on the Explore Diagram tab.

Comparison Diagrams

You can use comparison diagrams to compare two Nodes, Files or Cases and see which Project Items they do and do not have in common. The following project items can be displayed on a map:

- Nodes
- Files
- Cases
- Classifications
- Attributes
- Relationships
- Links

**Example: Create a Comparison Diagram of the 2 news articles to see what they have in common and how they differ.**

On the Explore tab, click on the Comparison Diagram button and choose to Compare Files. Then select the Files for both News Articles (in the news articles folder) and click OK. The graph will open up in the Detail View. You can simplify the graph by unticking elements such as Memo Links, Relationships and Files Coded. The graph below shows you which Nodes code both articles, and which do not.
Exporting Results and Project Items

The easiest way to export any Data Files, Nodes, Models, Query results and graphs is when they are displayed in **Detail View**. Click on the Detail View you want to export, and on the **Share** tab click **Export** (alternatively you can right-click within the Detail View and select Export). If you’re exporting **Text Search**, **Coding** or **Compound Queries**, choose what you want to export (Entire Content, Reference View or Summary View).

Choosing to export the **Entire Content** of a Node is useful for sharing your project findings (including audio and video clips) with colleagues who don’t have NVivo. The Node is exported as a series of web pages that are stored in your **My Documents** folder. You can open this file to access the 'mini website'. The Home page has the same name as the exported Node (for example, *adapting to climate change.html*), and it will have tabs containing the coded information in different Data File formats.

**Note:** If you want to move a Node HTML page, make sure you copy the folder containing the web page files to the new location as well. For Nodes that code audio and video Data Files, the coded sections of media files will be included upon export only if you coded directly on the audio or video timeline.

If you are exporting a **Matrix Coding Query**, a **Query Visualisation** (e.g., Word Cloud). Or list of items in the results of a Query, choose the file format for the exported file in the **Save As Type** list. In the **File Name** box, change the name if you like. Click **Browse** to select where you want to save it and click **OK**.

Exporting the Codebook

NVivo 12 has a new function where it can export your coding scheme as a Codebook. The Codebook includes the name of each Node and the description (if any) you have given them. This is **important** for reliability and validity purposes. You should always provide the Codebook to the readers of your research findings, so that they have a clear understanding of what each code means and the rules you used when applying codes to your data. With this information they can assess the validity and reliability of your results.

On the **Share** tab and click **Export Codebook** and click the Browse button for the **Save As** box. Choose the file type you want (Word or Excel), change the name if you need to and click **Save**. Then select the Nodes that you want to include in the Codebook and decide if you want
to include the number of Files and References as well by ticking the relevant box. Then click OK.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixed</td>
<td>A person's statement that they are indecisive as to whether their attitude is positive or negative. If a person states two different views, such as that they are positive about one aspect and negative about another, this is not mixed, but rather multiple, feelings.</td>
</tr>
<tr>
<td>Negative</td>
<td>A negative opinion or a disagreement where an opinion is not shared with someone or with a group.</td>
</tr>
<tr>
<td>Neutral</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>specifying positive opinion or agreement with other stakeholders or with an idea. A shared opinion</td>
</tr>
</tbody>
</table>

**Exporting Classification Sheets**

**Example: Exporting the Person classification summary sheet.**

If you want to export a spreadsheet containing your demographic or bibliographic information for example, you can export File or Case Classification Sheets as an Excel Workbook (*.xlsx). Open on the Classification Sheet you want to export (e.g., the Person classification in the Case Classifications folder). On the Share tab click Export. Select the appropriate export options: (e.g., export to Microsoft Excel). Choose a location to save the file and then click OK.

**Exporting Nodes**

In the List View, select the Node you want to export. On the Share tab, click Export and from the Export Options list select the appropriate option. Alternatively, you can just right click on the Detail View of an open Node and select Export. Click the Browse button under Save As and change the name/location of the exported files if you need to then click Save. Select the Properties and Related Content that you want to include then click OK.

**Exporting the List View of the Nodes Area**

You can export all the information shown in the List View of the Nodes Area into an Excel file. Right-click in the List View of the Node area, select Export List.

**Collaboration**

If you are collaborating with the team of researchers you can keep track of who coded what and how, by giving each team member accesses the project using their own ‘User Profile’. [You MUST use the same version of NVivo]. The easiest option for version control is for everyone to work on the same Project file at different times on a Standalone Project (a single file) logging-in with different user profiles.

If you want multiple people to work on a project at the same time, you can create a 'Master' project, and then give each person their own copy. You can import the work conducted by each person into the Master project selectively. On the Import tab, select Project and click the Browse button to find the Project copy you wish to import information from. In the Content Options area choose Selected (including content) and choose what you want to
import (e.g., Nodes and Coding). Decide how you want to deal with duplicate items (merge them or create new items) then click Import. Consider doing this at regular intervals.

**Note:** When you open the project, check the Current User Profile to make sure you are using your own. This is displayed on the Status Bar at the bottom of the NVivo window (with the person icon).

If you need to switch to your user profile without closing the Project: click the General tab and then under User enter the name and initials of your User Profile that you want to switch to.

**Interrater Reliability**

An important measure of the reliability of your coding scheme is to see how well multiple Coders (or Raters) agree with each other on the coding performed. You can conduct a Coding Comparison Query to get two measurements of inter-rater reliability: the percentage agreement (the number of units of agreement divided by the total units of measure within the Data File) and ‘Kappa coefficient’ (a statistic which compares your level of agreement to that which could be expected to occur by chance alone.

Checking interrater reliability is important to do at the beginning of the coding process, so that if any changes need to be made you don’t have to recode the whole Project. Find at least one other person familiar with your research and provide them with your Codebook. Give them access to your Project under a different User Profile. Select a small random sample of your Data and ask them to code this according to your Codebook. To make the process easier for the 2nd person, you can also limit the number of Nodes they need to code the data to, such as those that are particularly complex or potentially ambiguous. If you cannot find a 2nd person, code this sample of Data yourself a 2nd time after several weeks using a different User Profile. Then conduct an interrater reliability test to see how consistently you coded both times.

**Example: conduct a Coding Comparison Query between Effie and Henry.**

On the Explore tab click the Coding Comparison icon and select the users you wish to compare in the Compare coding between area (e.g., Effie Reeves and Henry Patterson). In the At box click Select and choose the Nodes you wish to include in the analysis (e.g., choose Nodes in selected sets, then click the Select button and choose the Nodes for coding comparison folder). Choose the specific Data Files to use in the Scope box (e.g. choose Data Files in selected sets, then click the Select button and choose the Interviews for coding comparison folder). Select the measure(s) you want to use: Kappa and/or Percentage Agreement. Click Run.

You can double-click on a row in the results where there is some disagreement (e.g., less than 80%) to compare the coding conducted and see where they disagreed for that File and Node. There will be a coloured coding stripe for Effie’s coding and another colour for Henry’s with their initials attached. The pink coding stripe represents the aggregated coding for that section (Effie and/or Henry’s coding).
Other resources

Training notes
To access training notes, visit the Research & learn webpage anulib.anu.edu.au/research-learn and select the skill area followed by the relevant course. You can register for a workshop and find other information.

Research & learn how-to guides
Explore and learn with the ANU Library’s how to guides (ql.anu.edu.au/howto). Topics covered are:

- Citations & abstracts
- Data Management
- E-books
- EndNote
- Finding books and more
- Finding journal articles and more
- Finding theses
- Increasing your research impact
- NVivo
- ORCID (Open Researcher and Contributor ID)
- Topic analysis
- Using Google scholar from off-campus

Subject guides
Find subject-specific guides (ql.anu.edu.au/subjectguides) and resources on broad range of disciplines. Such as:

- History, indigenous studies, linguistics and philosophy
- Criminal, human rights and taxation law
- Biochemistry and molecular biology, neurosciences and psychology
- Asia Pacific, Southeast Asia and East Asian studies
- Engineering
- Astronomy and astrophysics, earth sciences, mathematical sciences and natural hazards

Online learning
Online learning is available through ANU Pulse (ql.anu.edu.au/pulse), which can be accessed from both on and off campus by all ANU staff and students.

IT skills development modules available in ANU Pulse
- Microsoft Office (Access, Excel, OneNote, Outlook, PowerPoint, Project, Visio, Word)
- Microsoft Office (Mac)
- Adobe suite (Illustrator, Photoshop)
- Type IT

Training calendar
Select Events » near the bottom of the Library homepage to access our events calendar with upcoming training opportunities displayed day by day (ql.anu.edu.au/cal).

Feedback
Please provide feedback about today’s workshop via an online feedback form (ql.anu.edu.au/survey)