Working with Data

The City of White Rock aims to publish data in an open format and provide people with easy and reliable access to data, all while reducing barriers to entry, promoting innovation and eliminating unnecessary costs for businesses using the data.

Data is presented in different formats. This document contains information, tips and suggestions on how to use some of the formats presented in the Open Data Portal.

Structured Data

Structured data is usually in the form of text but follows a specific format giving it machine readability. Some more complex formats may be in binary format such as maps and GIS related files. You will find many different formats in the Open Data Portal. This section covers some of the popular formats and provides some tips on how to work with them.

CSV

The easiest method for working with a **Comma-Separated Values** (**CSV**) file is to use a spreadsheet application such as OpenOffice Calc or Microsoft Excel. Simply launch the application and open the .csv file to launch the application's CSV import wizard.

Character Encoding

One important detail is to pay attention to **Character Encoding** which can affect how the data, mainly the text is parsed during the import. This will be more noticeable when dealing with languages such as French which contain <u>extended characters</u>.

OpenOffice provides easy selection of the character encoding to be used. Figure 1 and Figure 2 illustrate an example where the downloaded CSV dataset file was encoded using a different format than the default UTF-8 for OpenOffice (illustration using NeoOffice).

00	Text Import – [orf	-ref-2012_03_31	-fra.csv]	
Import				
Character set	Unicode (UTE-8)		v	OK
				Cancel
Language	Default		•	
From row	1			Help
Separator options				
Fixed width				
 Separated by 				
🗹 Tab	Comma	Other		
Semicolon	Space			
Merge delimiters	3	Text delimiter	• •	
Other options				
Quoted field as t	ext			
Detect special n	umbers			
Column type				
Column type				
Standard			Standard	
1 Portefeuillemin	ist�riel 🦾		Titre l�gal	
2 Affairesautocht	coneset du d o veloppeme	n d u Nord canadi	enInstitutde la	
3 Affairesautocht	coneset du developpeme	ndu Nord canadi	enSoci�t��att	
4 Affaires trang	Prest Commerceinternat	tional	Centre de recl	
5 Affaires trang	Prest Commerceinterna	tional	Commissionde	
6 Affaires trang	Prest Commerceinterna	tional	Corporationco	
7 Affaires trang	Test Commerceinterna	tional	Exportationet	
8 Agence de promot	ion@conomiquedu Canada	aatlantique	Soci@t@@exr	
9 Agricultureet A	groalimentaire		Commissioncar	

Figure 1: Example of a CSV import wizard showing incorrect encoding

Screenshot showing the comma-delimited file import wizard in Neo Office. The file being imported contains French extended characters. The user has selected UTF-8 as the character encoding. The preview area is displaying the characters incorrectly as it is not the proper character encoding for this file.

Image: Second state statistic descendent of the statistic descendent of the statistic descendent of the statistic des en al statistic
Arial ▼ 10 ▼ B I I E E E E E E I I I E E E E I<
A1 ▼ f _i Σ Portefeuille minist riel 1 Portefeuille minist Titre l gal 2 Affaires autochtones et du d veloppement du Nord canadien Institut de la statistique des Pre 3 Affaires autochtones et du d veloppement du Nord canadien Soci t
1 Portefeuille ministôriel Titre lôgal 2 Affaires autochtones et du dôveloppement du Nord canadien Institut de la statistique des Pre 3 Affaires autochtones et du dôveloppement du Nord canadien Sociôtô dôattônuation des
1 Portefeuille ministoriel Titre logal 2 Affaires autochtones et du doveloppement du Nord canadien Institut de la statistique des Pressional des Affaires autochtones et du doveloppement du Nord canadien 3 Affaires autochtones et du doveloppement du Nord canadien Socioto do attonuation des
2 Affaires autochtones et du doveloppement du Nord canadien Institut de la statistique des Pro 3 Affaires autochtones et du doveloppement du Nord canadien Socioto do attonuation des
3 Affaires autochtones et du developpement du Nord canadien Societé de attenuation des
4 Affaires Itrang Pres et Commerce international Centre de recherches pour le c
5 Affaires Itrang res et Commerce international Commission de la capitale national
6 Affaires Itrang Pres et Commerce international Corporation commerciale cana
7 Affaires I Commerce international Exportation et developpemen

Figure 2: A CSV file imported using the incorrect character encoding

Accessible description of Figure 2

A comma-delimited file displayed in the Neo Office spreadsheet editor. The file contains French extended characters which were not properly imported due to using the incorrect character encoding.

Another common 8-bit character encoding in Canada is ISO–8859–1. In this scenario, selecting the ISO–8859–1 character encoding resolves the extended character issue as shown in Figure 3 and Figure 4.

Import OK Character set Western Europe (ISO-8859-1) Cancel Language Western Europe (DOS/OS2-863/French (Can.)) Cancel Western Europe (ISO-8859-1) Western Europe (ISO-8859-1) Help From row Western Europe (ISO-8859-11) Help Separator options Western Europe (ISO-8859-11) Help © Separator options Western Europe (ISO-8859-11/EURO) Western Europe (ISO-8859-11/EURO) © Separator options Semicolon Space Merge delimiters Text delimiter • Other options Space Standard Outoted field as text Detect special numbers B Fields B Count cancelien Institute It Affaireesatucchtoneset du développementuk Nord canadien Institute It Affaireeset commerce international Shiftaireeset commerce international Control candiensociété duté duté Affaireesétrangèreset Commerce international Comprationed Shiffaireesétrangèreset Commerce international Comprationed Shiffaireesétrangèreset Commerce international Corporationed Affaireesétrangèreset Commerce international Sporitationet Shiffairee	00				K.	
Character set Western Europe (ISO-8859-1) Cancel Uestern Europe (DOS/OS2-863/French (Can.)) Western Europe (DOS/OS2-863/French (Can.)) Western Europe (ISO-8859-11) Separator options Fixed width © Separated by © Tab © Comma © Other Space © Merge delimiters Fields Column type Standard Standard Fitresaturchtoneset du développementuk Nord canadien/Institut de 1e Affairesaturchtoneset Commerce international Contred e rect Affairesátrangèreset Commerce international Comportationed Affairesátrangèreset Commerce international Exportationet Affairesátrangèreset Commerce international Exportationet B Agence de promotionéconomiquedu Canada alantique Société dexpa	Import			\bigcirc		
Characteriset Western Europe (CSO-082-861/Icelandic) Language Western Europe (DOS/OS2-863/French (Can.)) Western Europe (DOS/OS2-863/French (Can.)) Western Europe (DOS/OS2-863/French (Can.)) Western Europe (ISO-8859-14) Western Europe (ISO-8859-14) Western Europe (ISO-8859-15/EURO) Western Europe (ISO-8859-15/EURO) Western Europe (ISO-8859-15/EURO) Western Europe (ISO-8859-15/EURO) Western Europe (ISO-8859-15/EURO) Western Europe (Windows-1252/WinLatin 1) ● Separated by ▼ Comma ● Separated by ♥ Comma ● Separated by ♥ Comma ● Separated by ♥ Comma ● Separated by ● Comma ● Separated by ● Comma ● Separated by ● Comma ● Outed field as text ● Detect special numbers Fields ● Standard ● Detect special numbers ● Standard ■ Standard ● Standard ● Standard	Character set	Mostern Europe (ICO 89	=======================================		OK	
Language Western Europe (DOS/OS2-863/Frenda (Can.)) Western Europe (DOS/OS2-865/Frenda (Can.)) Western Europe (SO-8859-14) Western Europe (SO-8859-14) Western Europe (SO-8859-14) Western Europe (SO-8859-15/EURO) Western Europe (SO-8859-15/EURO) Western Europe (Windows-1252/WinLatin 1)	Character set	Western Europe (ISO-88	69-1) 62.961/loglandia)			
Exinguing of the second second relation of the second second relation of the second	Language	Western Europe (DOS/OS	S2-863/Erenely (Can))		Cancel	
From row Western Europe (ISO-8859-14) Help Separator options Western Europe (ISO-8859-14) Western Europe (ISO-8859-14) Separator options Fixed width Western Europe (Windows-1252/WinLatin 1)	Lunguage	Western Europe (DOS/OS	S2-865(Nordic)			
Western Europe (ISO-8859-14) Western Europe (ISO-8859-15/EURO) Fixed width • Separated by Image: Separated by Image: Tab Image: Semicolon Semicolon Semicolon Space Image: Merge delimiters Tab Other options Quoted field as text Detect special numbers Fields B Column type Image: Standard Image: Standard Standard Standard Standard Image: Standard Image: Standard Standard The content of the state of t	From row	Western Europe (ISO-885	59-1)		Help	
Separator options Western Europe (ISO-8859-15/EURO) Western Europe (Windows-1252/WinLatin 1) Fixed width Separated by Tab Comma Other Semicolon Space Merge delimiters Text delimiter Text delimiter Other options Quoted field as text Detect special numbers Fields B Column type B Column type Standard Affaires autochtoneset du développement Nord canadien Institut de la 3 Affaires autochtoneset du développement Nord canadien Institut de la 4 Affaires étrangèreset Commerce international Common consissionde Affaires étrangèreset Commerce international Corporationez Réfaires étrangèreset Commerce international Société dexpa Affaires étrangèreset Commerce international Société dexpa Affaires étrangèreset Commerce international Société dexpa Agence de promotionéconomiquedu Canada atlantique Société dexpa Agricultureet Agroalimentaire 		Western Europe (ISO-885	59-14)			
Western Europe (Windows-1252/WinLatin 1) Fixed width Separated by Tab Comma Tab Comma Other Semicolon Space Merge delimiters Text delimiter Other options Quoted field as text Detect special numbers Fields Standard Standard Taires autochtoneset du développementu Nord canadien Institut de la Affaires autochtoneset du développementu Nord canadien Société datté Affaires étrangèreset Commerce international Centre de rect Stafiares étrangèreset Commerce international Commissionde Gaffaires étrangèreset Commerce international Comportatione Affaires étrangèreset Commerce international Société dexpa Affaires étrangèreset Commerce international Commissionde Affaires étrangèreset Commerce international Société dexpa Agence de pronotionéconomique Canada atlantique Société dexpa </td <td>Separator options</td> <td>Western Europe (ISO-885</td> <td>59-15/EURO)</td> <td></td> <td></td>	Separator options	Western Europe (ISO-885	59-15/EURO)			
Separated by Tab Comma Other Merge delimiters Text delimiter Other options Quoted field as text Detect special numbers Fields Standard Standard Fields Standard I Portefeuilleministériel B Column type B Standard I Portefeuilleministériel Fields B Column type B Column type B Column type B Column type Column type Column type Column type B Column type Co	Fixed width	Western Europe (Window	/s-1252/WinLatin 1)	<u> </u>		
Tab Comma Other Tab Comma Other Semicolon Space Merge delimiters Text delimiter Other options Quoted field as text Detect special numbers Fields B Column type B Standard Titre légal Affairesautochtoneset du développementul Nord canadien Institut de le Affairesautochtoneset du développementul Nord canadien Société datté Affairesétrangèreset Commerce international Commissionde Affairesétrangèreset Commerce international Commissionde Affairesétrangèreset Commerce international Exportationet Agence de promotionéconomiquedu Canada atlantique Société dexpa Agence de promotionéconomiquedu Canada atlantique	Separated by					
✓ Tab ✓ Comma Other ○ Semicolon Space ○ Merge delimiters Text delimiter ♥ ○ Other options ● ○ Quoted field as text ● ● ○ Detect special numbers ● Fields ● ✓ Column type ● ✓ Standard Standard ↑ Portefeuilleministériel ● ✓ Affaires autochtoneset du développementdu Nord canadien Institut de 1a ③ Affaires autochtoneset du développementdu Nord canadien Institut de 1a ③ Affaires étrangèreset Commerce international Centre de recl ⑤ Affaires étrangèreset Commerce international Commissionde ⑥ Affaires étrangèreset Commerce international Corporationcc ⑦ Affaires étrangèreset Commerce international Exportationet ⑧ Agence de promotionéconomiguedu Canada atlantique Société dexpa ⑧ Agricultureet Agroalimentaire Commissioncar	Copurated by					
Semicolon Space Merge delimiters Text delimiter Other options	🗹 Tab	🗹 Comma	Other			
Merge delimiters Image: Merge delimiters Other options Quoted field as text Detect special numbers Fields Column type Standard 1 Portefeuilleministériel 2 Affaires autochtoneset du développementdu Nord canadien Institut de la 3 Affaires autochtoneset du développementdu Nord canadien Institut de la 3 Affaires autochtoneset du développement Nord canadien Société datté 4 Affaires étrangèreset Commerce international Commissionde 6 Affaires étrangèreset Commerce international Comportationed 7 Affaires étrangèreset Commerce international Comportationed 8 Agence de promotionéconomique Canada atlantique 9 Agricultureet Agroalimentaire	Semicolon	Space				
Other options Quoted field as text Detect special numbers Fields Column type Standard Portefeuilleministériel Affaires autochtoneset du développementdu Nord canadien Institut de la Affaires autochtoneset du développementdu Nord canadien Société datté Affaires autochtoneset du développementdu Nord canadien Société datté Affaires étrangèreset Commerce international Centre de recl Affaires étrangèreset Commerce international Comporationce Affaires étrangèreset Commerce international Exportationet Agence de promotionéconomiguedu Canada atlantique Société dexpa Agricultureet Agroalimentaire	Merge delimiters		Text delimiter	"		
Other options Quoted field as text Detect special numbers Fields Column type Standard Portefeuilleministériel Affaires autochtoneset du développementdu Nord canadien Institut de la Affaires autochtoneset du développementdu Nord canadien Institut de la Affaires étrangèreset Commerce international Centre de reci Affaires étrangèreset Commerce international Commissionde Affaires étrangèreset Commerce international Commissionde Affaires étrangèreset Commerce international Comportioncc Affaires étrangèreset Commerce international Comportioncc Affaires étrangèreset Commerce international Exportationet Agence de promotionéconomiquedu Canada atlantique Société dexpa Agence de promotionéconomiquedu Canada atlantique Commissioncar						
Quoted field as text Detect special numbers Fields Column type Standard Portefeuilleministériel Affaires autochtoneset du développementdu Nord canadien Institut de la Affaires autochtoneset du développementdu Nord canadien Société datté Affaires étrangèreset Commerce international Centre de recd S Affaires étrangèreset Commerce international Commissionde Affaires étrangèreset Commerce international Commissionde Affaires étrangèreset Commerce international Comportationet Affaires étrangèreset Commerce international Commissionde Affaires	Other options					
Detect special numbers Fields Column type Standard Portefeuilleministériel Affaires autochtoneset du développementdu Nord canadien Institut de la Affaires autochtoneset du développementdu Nord canadien Institut de la Affaires autochtoneset du développementdu Nord canadien Société datté Affaires étrangèreset Commerce international Centre de rece Affaires étrangèreset Commerce international Commissionde Affaires étrangèreset Commerce international Exportationet Agence de promotionéconomiquedu Canada atlantique Société dexpa Agricultureet Agroalimentaire	Quoted field as text					
Fields Column type Standard B Standard 1 Portefeuilleministériel Titre légal 2 Affaires autochtoneset du développement Nord canadien Institut de la 3 Affaires autochtoneset du développement Nord canadien Société datté 4 Affaires étrangèreset Commerce international 5 Affaires étrangèreset Commerce international 6 Affaires étrangèreset Commerce international 7 Affaires étrangèreset Commerce international 8 Agence de promotion économique Canada atlantique 9 Agricultureet Agroalimentaire	Detect special put					
Fields Column type Standard Standard 1 Portefeuilleministériel Titre légal 2 Affaires autochtoneset du développement du Nord canadien Institut de la 3 Affaires autochtoneset du développement du Nord canadien Société datté 4 Affaires étrangèreset Commerce international Centre de reck 5 Affaires étrangèreset Commerce international Commission de 6 Affaires étrangèreset Commerce international Corporation co 7 Affaires étrangèreset Commerce international Exportation et 8 Agence de promotion économi que du Canada atlantique Société dexpa 9 Agricultureet Agroalimentaire Commission car	Detect special nu	mbers				
Standard Standard 1 Portefeuilleministériel Titre légal 2 Affaires autochtoneset du développementdu Nord canadien Institut de la 3 Affaires autochtoneset du développementdu Nord canadien Société datté 4 Affaires étrangèreset Commerce international 5 Affaires étrangèreset Commerce international 6 Affaires étrangèreset Commerce international 7 Affaires étrangèreset Commerce international 8 Agence de promotionéconomiquedu Canada atlantique 9 Agricultureet Agroalimentaire	Fields		\frown			
Standard Standard 1 Portefeuilleministériel Titre légal 2 Affaires autochtoneset du développementdu Nord canadien Institut de la 3 Affaires autochtoneset du développementdu Nord canadien Société datté 4 Affaires étrangèreset Commerce international 5 Affaires étrangèreset Commerce international 6 Affaires étrangèreset Commerce international 7 Affaires étrangèreset Commerce international 8 Agence de promotionéconomiquedu Canada atlantique 9 Agricultureet Agroalimentaire	0-1		В			
StandardStandard1PortefeuilleministérielTitre légal2Affaires autochtoneset du développementdu Nord canadien Institut de la3Affaires autochtoneset du développementdu Nord canadien Société datté4Affaires étrangèreset Commerce internationalCentre de recl5Affaires étrangèreset Commerce internationalCommissionde6Affaires étrangèreset Commerce internationalCorporationce7Affaires étrangèreset Commerce internationalExportationet8Agence de promotionéconomiquedu Canada atlantiqueSociété dexpa9Agricultureet AgroalimentaireCommissioncar	Column type	· · · · · · · · · · · · · · · · · · ·				
1PortefeuilleministérielTitre légal2Affaires autochtoneset du développementdu Nord canadien Institut de la3Affaires autochtoneset du développementdu Nord canadien Société datté4Affaires étrangèreset Commerce internationalCentre de recl5Affaires étrangèreset Commerce internationalCommissionde6Affaires étrangèreset Commerce internationalCorporationco7Affaires étrangèreset Commerce internationalExportationet8Agence de promotionéconomique Canada atlantiqueSociété dexpa9Agricultureet AgroalimentaireCommissioncar	Standard			Standard		
2Affaires autochtoneset du développementdu Nord canadien Institut de la3Affaires autochtoneset du développementdu Nord canadien Société datté4Affaires étrangèreset Commerce internationalCentre de recl5Affaires étrangèreset Commerce internationalCommissionde6Affaires étrangèreset Commerce internationalCorporationco7Affaires étrangèreset Commerce internationalExportationet8Agence de promotion économique uCanada atlantiqueSociété dexpa9Agriculture AgroalimentaireCommissioncar	1 Portefeuillemini	stériel		Titre légal		
3Affaires autochtoneset du développement du Nord canadien Société datté4Affaires étrangèreset Commerce internationalCentre de reck5Affaires étrangèreset Commerce internationalCommission de6Affaires étrangèreset Commerce internationalCorporation cc7Affaires étrangèreset Commerce internationalExportation et8Agence de promotion économique du Canada atlantiqueSociété dexpa9Agriculture AgroalimentaireCommission car	2 Affairesautochto	oneset du développeme	entdu Nord canadien	Institutde la		
4Affairesétrangèreset Commerce internationalCentre de recl5Affairesétrangèreset Commerce internationalCommissionde6Affairesétrangèreset Commerce internationalCorporationco7Affairesétrangèreset Commerce internationalExportationet8Agence de promotionéconomiquedu Canada atlantiqueSociété dexpa9Agricultureet AgroalimentaireCommissioncar	3 Affairesautochto	oneset du développeme	entdu Nord canadien	Société datté		
5Affairesétrangèreset CommerceinternationalCommissionde6Affairesétrangèreset CommerceinternationalCorporationco7Affairesétrangèreset CommerceinternationalExportationet8Agence de promotionéconomiquedu Canada atlantiqueSociété dexpa9Agricultureet AgroalimentaireCommissioncar	4 Affairesétrangèreset Commerceinternational Centre de recl					
6Affairesétrangèreset Commerce internationalCorporationco7Affairesétrangèreset Commerce internationalExportationet8Agence de promotioné conomique du Canada atlantiqueSociété dexpa9Agriculture AgroalimentaireCommissioncar	5 Affairesétrangèreset Commerceinternational Commissionde					
7Affairesétrangèreset CommerceinternationalExportationet8Agence de promotionéconomiquedu Canada atlantiqueSociété dexpa9Agricultureet AgroalimentaireCommissioncar	6 Affairesétrangèreset Commerceinternational Corporationco					
8 Agence de promotionéconomiquedu Canada atlantique Société dexpa 9 Agriculture Agroalimentaire Commissioncar	7 Affairesétrangèreset Commerceinternational Exportationet					
9 Agricultureet Agroalimentaire Commissioncar	8 Agence de promotionéconomiguedu Canada atlantigue Société dexpa			Société dexpa		
	9 Agricultureet Agroalimentaire Commissioncar					

Figure 3: A CSV import wizard showing proper character encoding

Screenshot showing the comma-delimited file import wizard in Neo Office. The file being imported contains French extended characters. The user has selected the appropriate ISO-8859-1 encoding and the preview area is displaying the characters properly.

000					
	i • 🟠 📃 🗮 🗾 🔝 🖨 🖄 🗖 🔜	🕺 🗅 🖆 🖤 I Q			
	Arial • 10 • B / <u>U</u> = = =				
A1	• $f_{(x)}$ Σ = Portefeuille ministér	iel			
	Α 🦯				
1	Portefeuille ministériel	Titre légal			
2	Affaires autochtones et du développement du Nord canadien	Institut de la statistique des Prem			
3	Affaires autochtones et du développement du Nord canadien	Société datténuation des répercu			
4	Affaires étrangères et Commerce international	Centre de recherches pour le dév			
5	Affaires étrangères et Commerce international	Commission de la capitale nation			
6	Affaires étrangères et Commerce international	Corporation commerciale canadie			
7	Affaires étrangères et Commerce international	Exportation et développement Ca			

Figure 4: A CSV file imported into NeoOffice using the proper character encoding

Accessible description of Figure 4

A comma-delimited file displayed in the Neo Office spreadsheet editor. The file contains French extended characters which were properly imported using the correct character encoding.

You may notice character encoding issues when opening a CSV file in Microsoft Excel as shown in Figure 5. For this example, we will be using Microsoft Excel in Microsoft Office Professional Plus 2010.

	Α	В
1	language	title
2	English	RISO (CFS St. John's) - Fresh Fruit and Vegetables
3	French	OCIR (MDN St. John's) - Produits de fruits et légumes frais
4	English	Software Licensing Supply Arrangement
5	French	Arrangements en matière d'approvisionnement portent sur l'achat de licences d
6	English	FREESTANDING FURNITURE
7	French	MEUBLES AUTOSTABLES
8	English	FREESTANDING FURNITURE
9	French	MEUBLES AUTOSTABLES
10	English	Industry Engagement Day Notice - Sep 24 2013 - Converged Communications Servi
11	French	Avis relatif à la Journée de consultation de l'industrie - le 24 sep 2013 - Service
12	English	R&O - Leopard 2 Turret Electro/Optic Components
13	French	R&R - Pièces électro-optiques de la tourelle du Leopard 2
14	English	CSC- PORT CARTIER - CHEESE

Figure 5 - File opened in Excel using incorrect file encoding

Accessible description of Figure 5

A screenshot of a comma-delimited file opened in Excel showing incorrectly decoded French characters.

The character encoding can be manually selected using the data import wizard. The first step is to start with a new Excel workbook. Go to the Data tab and click on From Text which will launch the import wizard (Figure 6).

The first step in the wizard is to select the appropriate encoding. In this example, we are selecting the UTF-8 encoding which is located near the bottom of the list. Once selected, you can preview the character decoding in the data preview pane of the wizard.

Text Import Wizard - Step 1 of 3		? <mark>x</mark>	
The Text Wizard has determined that your data is Delimited.			
If this is correct, choose Next, or choose the data typ	e that best describes your data.		
Original data type			
Choose the file type that best describes your data:			
Delimited - Characters such as commas of the second sec	r tabs separate each field.		
Fixed width - Fields are aligned in columns with the second se	with spaces between each field.		
5			
Start import at <u>r</u> ow: 1 File <u>o</u> rigin:	Windows (ANSI)	-	
	10017 : Ukrainian (Mac)		
	65000 : Unicode (UTF-7)		
	20127 : US-ASCII		
Preview of file	1258 : Vietnamese (Windows)		
	20005 : Wang Taiwan	· · · · · · · · · · · · · · · · · · ·	
1 "language", "title", "reference_number", "solicitation_number", "amendment_ni *			
3 "French", "OCIR (MDN St. John's) - Produits de fruits et lÃ@gumes frais",			
4 "English", "Software Licensing Supply Arrangement", "PW-\$\$EE-015-22221", "EN			
5 "French", "Arrangements en matiÀ"re d'approvisionnement portant sur l'acha 🔻			
<		•	
C	ancel < Back <u>N</u> ext >	Einish	

Figure 6 - Selecting UTF-8 in Text Import Wizard

Accessible description of Figure 6

A screenshot showing Step 1 of 3 of the Text Import Wizard in Microsoft Excel. The file origin list box is open with Unicode UTF-8 selected.

The next step (Figure 7) allows you to specify the proper delimiter(s) used in the data file. Once again, the Data preview pane allows you to preview your data and see if it is parsing the data correctly.

Text Import Wizard - Step 2 of 3
This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.
Delimiters
Tab
Semicolon Treat consecutive delimiters as one
Comma Taxt qualifier:
Space
Other:
Data <u>p</u> review
language title
English RISO (CFS St. John's) - Fresh Fruit and Vegetables French DCIR (MDN St. John's) - Produits de fruits et lÂ@gumes frais
English Software Licensing Supply Arrangement
French Arrangements en matiÃïre d'approvisionnement portant sur l'achat 🗸 🔻
Cancel < <u>B</u> ack <u>N</u> ext > <u>F</u> inish

Figure 7 - Selecting delimiters in Text Import Wizard

A screenshot showing Step 2 of 3 of the Text Import Wizard in Microsoft Excel. The Comma delimiter is checked and the data preview pane is showing the data separated into columns.

Optionally, you can specify the data types for each of the columns (Figure 8). In most cases, simply clicking Finish will suffice.

Text Import Wizard - Step 3 of 3	3			
This screen lets you select each column and set the Data Format. Column data format General Text Date: DMY Do not import column (skip)				
General General language title Snglish RISO (CFS St. John's) - Fresh Fruit and Vegetables OCIR (MDN St. John's) - Produits de fruits et lÃ@gumes frais Snglish Software Licensing Supply Arrangement French Arrangements en matiÃ"re d'approvisionnement portant sur l'achat (* Image: Cancel < Back				

Figure 8 - Selecting data formats in Text Import Wizard

A screenshot showing Step 3 of 3 of the Text Import Wizard in Excel. This step has settings allows the user to set the data formats for each of the columns. The defaults are used for this example.

Figure 9 shows the characters properly decoded in Excel having used the Text Import Wizard.

	А	В	
1	language	title	
2	English	RISO (CFS St. John's) - Fresh Fruit and Vegetables	
3	French	OCIR (MDN St. John's) - Produits de fruits et légumes frais	
4	English	Software Licensing Supply Arrangement	
5	French	Arrangements en matière d'approvisionnement portagt sur l'achat de licences de	
6	English	FREESTANDING FURNITURE	
7	French	MEUBLES AUTOSTABLES	
8	English	FREESTANDING FURNITURE	
9	French	MEUBLES AUTOSTABLES	
10	English	Industry Engagement Day Notice - Sep 24 2013 - Converged Communications Servi	
11	French	Avis relatif à la Journée de consultation de l'industrie - le 24 sep 2013 - Services de	
12	English	R&O - Leopard 2 Turret Electro/Optic Components	
13	French	R&R - Pièces électro-optiques de la tourelle du Leopard 2	
<mark>14</mark> Figu	14 English CSC- PORT CARTIER - CHEESE Figure 9 - Data imported into Excel using proper character encoding		

A screenshot showing a comma-delimited file imported into Excel using the Text Import Wizard. The French characters are shown properly having applied the appropriate character encoding.

Byte Order Mark (BOM)

You may discover an interesting special byte at the beginning of a file. This byte is called the Byte Order Mark or BOM. This marker is used to signal the size of the character encoding used in the UTF encoded file. For example, UTF–16 will render as U+FEFF (Figure 5) or $b\ddot{y}$ in a text editor capable of displaying this value. In general, you do not need to concern yourself with the BOM as it is used by the application you are using. However, it can be a useful byte to help you select the correct file encoding during an import if you are having issues.

000	1. less
<pre><u+feff>Indicator_Name,Indicator_Code,[Ht _Time].[\ar].&[YR2000],[HNP_Ctry].[List] &[YR2011],[NNP_Ctry].[List].&[LIC].[HNP_' _Ctry].[List].&[LIC].[HNP_Time].[Year].& &[SAS].[HNP_Time].[Year].&[YR1990],[HNP_C me].[Year].&[YR2011],[HNP_Ctry].[List].& "Malnutrition prevalence, weight for age ,42.67046,33.88734,"</u+feff></pre>	NP_Ctry].[List].&[AFG].[HNP_Time].].&[AFG].[HNP_Time].[Year].&[YR201 Time].[Year].&[YR1990],[HNP_Ctry]. [YR2010],[HNP_Ctry].[List].&[LIC]. Ctry].[List].&[SAS].[HNP_Time].[Ye [SAS].[HNP_Time].[Year].&[YR2011] (% of children under 5)",SH.STA.M.
"Malnutrition prevalence, height for age 1,50.58797,39.7836,	(% of children under 5)",SH.STA.S
Prevalence of wasting (% of children unde Low-birthweight babies (% of births),SH.S Exclusive breastfeeding (% of children un Vitamin A supplementation coverage rate of 860490001,85.1644867893504,,,50.22503874 Prevalence of overweight (% of children un 72,	er 5),SH.STA.WAST.ZS,,,,,11.43469, STA.BRTW.ZS,,,,,SH.STA.BFED.ZS,,,,, nder 6 months),SH.STA.BFED.ZS,,,,, (% of children ages 6-59 months),S 718,73.1598282328674 under 5),SH.STA.0WGH.ZS,,,,,1.4228
Prevalence of undernourishment (% of popul ,21.4592615854612,18.4848502097362,17.788 Consumption of iodized salt (% of househousehousehousehousehousehousehouse	ulation),SN.ITK.DEFC.ZS,,,,,,33.04 34326669054 olds),SN.ITK.SALT.ZS,,,,,,,,,,,,

Figure 10: A Byte Order Mark in a data file

Accessible description of Figure 10

A Byte Order Mark displayed in a text editor.

JSON

JavaScript Object Notation (JSON) is a structured open data format that is lightweight and very well suited for mobile applications. It is used either natively or as a library available in most web and mobile development languages. Detailed information on JSON can be found on the <u>official</u> JSON web site. Here is a simple example of JSON data:

```
1. "metadata":{
2. "request":{
3. "dateCreated":"2014-01-01T00:00:00+00:00",
4. "name": "Example JSON Dataset" },
5. "items":[
6. { "name":"Item 1", "colour":"red" },
7. { "name":"Item 2", "colour":"blue" }
8. ]
9. }
```

We have lots of great examples of working with JSON data and RESTful APIs in various programing languages.

XML

XML (eXtensible Markup Language) defines a set of rules for encoding documents in machinereadable form. It is defined in the <u>XML 1.0 Specification</u>.

Geo Formats

GeoTIFF

GeoTIFF files are similar to standard graphical TIFF files; they are viewable using a regular image viewer that supports the TIFF format. However, GeoTIFF files contain extra metadata related to the geographical nature of the image. Refer to the official sites on GeoTIFF for additional information on the format:

- Official GeoTIFF web site
- <u>GeoTIFF Specification</u>

An open source library exists to enable your program the ability to extract the metadata information from a GeoTIFF. Below is an example on how to install and run a simple C program to extract metadata information.

Here we are using the Homebrew package manager on OSX to install the GeoTIFF library and use a utility to read a GeoTIFF:

```
1. $ brew install libgeotiff
```

```
2. $ listgeo <filename>.tif
```

Once the GeoTIFF library is installed, you should be able to compile a simple C program to read the metadata:

```
1. /**
2. * Simple example of a C program to read metadata from a GeoTIFF.
3.
4. * NOTE: This is a simple example with very little error handling.
5.
6. * @license http://data.gc.ca/eng/open-government-licence-canada
7. */
8. #include <stdio.h>
9. #include <stdlib.h>
10. #include <xtiffio.h>
11. #include <geotiffio.h>
12. #include <geokeys.h>
13.
14. int main(void) {
15.
        TIFF *tif = (TIFF*)0; // TIFF-level descriptor
16.
      GTIF *gtif = (GTIF*)0; // GeoKey-level descriptor
17.
18.
       double major axis key = 0;
19.
20. // Open TIFF file
```

```
21.
        tif = XTIFFOpen("geo.tif", "r");
22.
        if (!tif) {
            printf("Unable to open GeoTIFF file.\n");
23.
24.
            exit(-1);
        }
25.
26.
27.
       // Open GTIF Key parser; keys will be read at this time.
28.
        gtif = GTIFNew(tif);
29.
       if (!gtif) {
30.
            printf("Unable to access GTIF metadata.\n");
31.
            exit(-1);
32.
        }
33.
34.
       // Read a GTIF Key pair
35.
       if (GTIFKeyGet(gtif, GeogSemiMajorAxisGeoKey, &major_axis_key, 0,
 1)) {
36.
            printf("GeogSemiMajorAxisGeoKey: %f\n", major_axis_key);
37.
        } else {
38.
            printf("Unable to locate GeogSemiMajorAxisGeoKey\n");
        }
39.
40.
41.
       // Clean up and exit.
42.
       GTIFFree(qtif);
43.
      XTIFFClose(tif);
44.
       return 1;
45. }
```

Output:

```
1. $ ./geotiffmeta
```

2. GeogSemiMajorAxisGeoKey: 6378137.000000

Shapefiles

Browse Shapefile Datasets

ESRI Shapefiles (SHP) are popular geospatial vector data files which are accompanied by a series of files necessary to use the SHP file. As listed on Wikipedia, the following is usually packaged into a compressed ZIP file:

Mandatory files:

- .shp shape format; the feature geometry itself
- .shx shape index format; a positional index of the feature geometry to allow seeking forwards and backwards quickly
- .dbf attribute format; columnar attributes for each shape, in dBase IV format

Optional files:

• .prj — projection format; the coordinate system and projection information, a plain text file describing the projection using well-known text format

- .sbn and .sbx a spatial index of the features
- .fbn and .fbx a spatial index of the features for shapefiles that are read-only
- .ain and .aih an attribute index of the active fields in a table
- .ixs a geocoding index for read-write shapefiles
- .mxs a geocoding index for read-write shapefiles (ODB format)
- .atx an attribute index for the .dbf file in the form of shapefile.columnname.atx (ArcGIS 8 and later)
- .shp.xml geospatial metadata in XML format, such as ISO 19115 or other XML schema
- .cpg used to specify the code page (only for .dbf) for identifying the character encoding to be used

Source: Shapefile - Wikipedia.org

Links to useful tools:

- Online Shapefile Dump Utility
- <u>Perl Shapefile Extension</u>
- <u>Python Shapefile Library</u>
- <u>C Shapefile Library</u>

Converting Shapefiles to GeoJSON

There is a great write-up by Ben Balter on <u>converting Shapefiles to GeoJSON</u> using gdal's ogr2ogr utility as follows:

1. ogr2ogr -f GeoJSON -t_srs crs:84 [name].geojson [name].shp