Local Open Data

The Connemara Open Data Project

Connemara Open Data Project
August 2017

Contents

Connemara Open Data Project

Introduction

Rural and small urban areas will increasingly be disadvantaged if they do not create local open data resources to support their individuals, businesses, administrators and policy makers.

In 2011 the Connemara Programme started a project to document and digitise the data and information about the Connemara area in Ireland.
The purpose of this project has been to provide a locally based, open data resource for individuals, businesses, organisations and administrators. It can be used for activities such as marketing, communications, web and app development and decision making processes. Alignment and integration of many of data sets to county, national and EU open data resources is presently being done as Phase 2.

These local data sets help to reduce the cost of content and provides a consistent and rich set of data for all businesses and organisations in Connemara. They can also service the needs of external businesses and administrators and content users who are interested in representing Connemara data to a wider audience.

Low / No cost tools have been used in the processes followed during the project. This was done to explore how much could be achieved without incurring large set up and operational overheads. Its success as a strategy removes many of the cost concerns other rural areas and small urban communities may have in approaching open data projects.

11 levels of data and information have been examined and digitised. There are: Place, Business, Community, Economy, Social, Topography, Infrastructure, Architecture, Culture, Skills, Flora and Fauna. These provide a detailed 360-degree data view of Connemara.

Over 25,000 data points are presently available along with over 3000 up to date photographs.

The available data was codified and normalised and is presented from two perspectives:

1. Data linked to Place: (Provides a place data resource for local users: EG Clifden Businesses)
2. Subject specific: (Provides specific data sets against a specialised need: EG Fishing Locations)

Phase 1 access to the data sets is through a simple web site (www.connemaraprogramme.com/opendata). This is to immediately open the data to the local community. Phase 2 will establish a relationship with an open data repository to assist with integrating the data sets to the County and National open data repositories.

Presently users can download excel (.XLSX) format files for the data sets. Graphic resources are provided in JPG, GIF or PNG format in a ZIP file. Graphics are available for the individual places and also for subject specific data sets.

The .XLSX format for the data sets was chosen so that most businesses and organisations can make use of the data immediately without acquiring new skills. The Graphics are usually 400 x 300 pixels in size and are highly compressed in order to be web usable without further user tailoring.

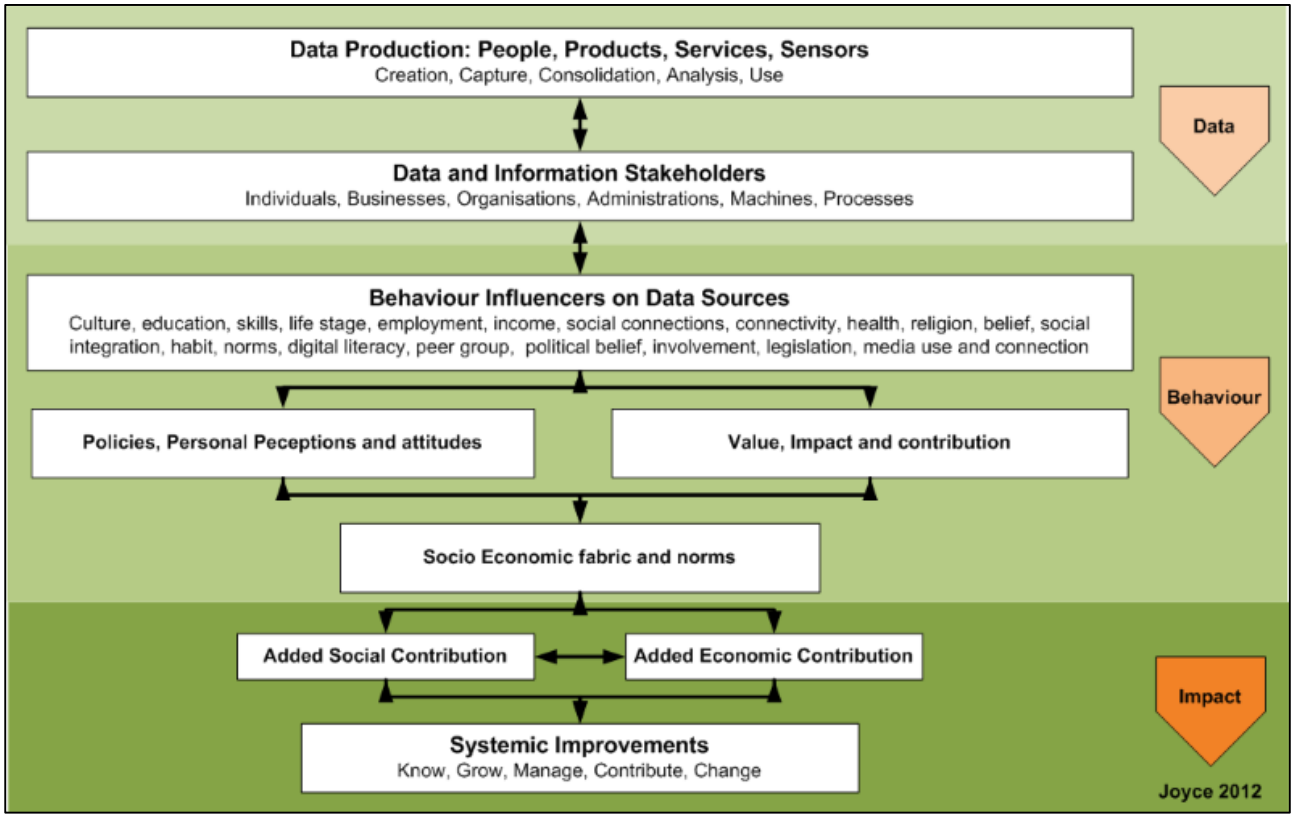
English is the standard language. Text elements have been controlled for colloquialisms, Irish specific character sets and non-standard linguistic constructs. This has been done to ensure a quality and consistency of input to automatic translation capabilities which can support the linguist needs of non-English speakers in the community and as tourists.

The following sections provide an overview of the processes, procedures and tools used in managing the Connemara Open Data Project.

Overview Diagram

A model of the underlying dynamics of the data environment of Connemara was created. This was done to assist with identifying, prioritising and managing the various actors, influences and stakeholders in the data environment.

It has also assisted in explaining to various stakeholders how open data is a resource and has a positive impact on the entire community now and in the future.



It also helped with the design of a data audit and the easier identification of core data sets for the project.

The Initial data availability Audit

This audit is an examination of the existing potential data resources, what and where they are, their priority and if they can be immediately used or what additional processing to be of use to the community. It also identifies missing or incomplete data sets that would be of value.

Does the data exist?

Yes
Document the data source, content and who it is of value to. Prioritise it as to its potential contribution to an actor or actors in the area and any information that is available on it.

NO
Decide the priority of the missing data set. Is it worth creating it? Does it have immediate value to an actor or group of actors in the area? How important is it overall?

Is data in digital form?

Yes
Document the digital format that the data is in. Do you have the tools to integrate / transcribe it to a database or the format you use to collect and store data in?

NO
Decide on the priority of the data set and how much effort is required to transcribe it to a digital format. Do you have to tools to extract it or if not how much manual effort is required? Will it contribute positively to the project and its users?

Is it publicly available?

Yes
Document the owner and source, permissions and ability to reuse the data.

NO
Locate the owner of the data or information source and request permission to reuse it as an open source resource. It is useful to explain why the data is important to the project

Is the data available for free?

Yes
Note that the data is free in your audit

NO
Contact the data owner and discuss terms of use / cost. Often a sub set of the data can be negotiated for free in return a service or “branding” trade off.

Is the data available online?

Yes
Note the URL and any logins and Passwords you require to access the data.

NO
Note the source of the data and where it can be obtained. Prioritise data sets that you wish to source as hard copy.

Is the data machine readable?

Yes
What is the format? Decide how you will transcribe the data to your data base or storage medium.

NO
How important is the data to the project? What is its value? If sourced can it the digitised with the tools you have available?

Available in bulk

Yes
Examine the data set to see if it is comprehensive enough to fulfil the needs of the project. Can it be augmented / complimented by another data set?

NO
Examine the resources required to capture the individual data records. Is it worth the effort?

Is the Data Openly licensed?

Yes
Document the licence used and retrieve the data.

NO
Contact the data owner and enquire if the data set or a useful data set could be permitted for use under an open license.

Is the data provided on a timely and up-to-date basis?

Yes
Ensure that a creation and update parameter is available as part of the metadata set. The date of availability of the data should also be collected.

NO
Locate a creation and update parameter if available. Enquire as to the frequency of update from the data owner. Is it timely enough to fulfil the project goals? What effort will be required to maintain the data set?

The audit helped identify the core data sets that could be sourced rapidly, cost effectively and efficiently.

Core Local Open Data Sets

The core data sets of the Connemara Open Data Project (CODP) are designed to be additional and complimentary to PSI (Public Service Information) data sets that may be available from Irish public administration sources. The census 2016 provided a vast amount of data sets which were integrated to many of the location data sets.

Many of the local data sets exist and can be extracted from non-copyright hard copy resources. Others require the creation of data set records. (see: *Data Sources and Tools* below)

The data sets were chosen for their relevance to a location or area by reference to its economic structure *EG: areas with a high degree of tourist activity will benefit most from a different set of core data sets than an agriculture focused / non-tourist area.*

Data Set	Characteristics
Activities	Activity resources available in the community or area
Architecture	Captures the key human constructed features of an area
Communities	Captures details of individual communities / towns / villages
Community Facilities	Captures the structural elements of a community
Community Services	Captures voluntary services and supports in a community
Culture	Captures the key intangible cultural artefacts (dance music etc)
Economic actors	Businesses and organisations with an economic purpose
Economic Structure	Basic information on the economic structure of the area
Heritage	Captures customs, traditions, traditional skills of an area
History	Captures the key past events / activities in an area
Landscape	Captures the topographical features of an area
Transport	Elements of local transport not available elsewhere (distances)
Weather	Captures to meteorological characteristics of an area
Wildlife	Captures the flora and fauna of an area

Many other data sets can be created to clearly reflect the unique social, economic, cultural, historical and physical characteristics of a community or area.

These data provide a rich, free, digital content resource for Individuals, businesses, organisations and administrators.

They cut the costs of digital platform set up, increase consistency of information to end users and can, and do, encourage further content input from local residents.

For each of these core data sets a single set of data fields emerged as being adequate to represent the data sets.

Key Data Fields

The fields listed below have proven to give the flexibility and richness required to capture, categorise and quantify the data sets listed above.

Text Data Fields	**Purpose**
Main Category	Provides a top level sort criterion
Sub Category / Sub Categories	Provides added detail under the top level category
Name	Unique name (descriptor) for an item being recorded
Place	The community to which the record is attributed / linked
County / region / State	The administrative level in which the Place is located
Country	The nation within which the place is located
Latitude	A latitude coordinate for mapping use
Longitude	A longitude coordinate for mapping use
Historical Period	A bulk period relevant to the record (e.g. 1800 – 1900)
Historical Date	A specific year relevant to the record (if available)
Data	The relevant datum or data content of the record
Comments	Any text based comments required to explain the record
Source	The source of the data record
Web Address	A URL that links to the data or data set (if available)
License	The open data license used
Creation Date	The date on which the record was created
Graphics Data Fields	**Purpose**
Graphic name	A unique descriptive identifier for a graphic (e.g. xyz castle001)
Graphic Web Link	Link to the graphic
Graphic description	What the graphic represents and its format
Source	The source of the data record
License	The open data license used
Creation Date	The date on which the record was created

Consistent categorisation of the data within a data set is critical to ensuring that the records are codified and usable in a logical fashion. The use of one or more sub category is useful to providing increased flexibility of use, sorting and extraction.

English is the standard language of the Connemara Open Data Project. The use of a single language in the data base makes it easier to manage.

All text elements should avoid localisms and colloquialisms.

The use of formal linguistic constructs, consistent spelling and grammar rules allows for easier and more meaningful auto translation by online tools such as Google Translate and other emerging auto translate tools. This is especially important in tourism dependent areas.

Metadata

In Phase 1 of the Connemara Open data project metadata was not a priority as provision of data to a local user base was more important and it would have been of little use to them.

For Phase 2 which looks to integrate the data sets to a local, national or EU open data repository the creation of metadata for the data sets is critical as this data describes the data set, its format, where it is found, license and other information useful to those looking to reuse it.

The set of variables below are the key metadata elements that are recommended or required for any data set being presented to a formal open data resource run by county, regional, national or EU.

Type of Metadata	Purpose
Dataset	The name of the data set including the type (xxxx.csv)
Description	A detailed description of what the file contains and its purpose and creation
Publisher	Who is publishing this file
Title	The title to appear when someone accesses the open data portal
Access URL	The URL of the data file (full path)
Issued	The date when the file was made live on the server
Language	What language is used in the file (e.g. English)
License	What open license is used for this file and a link to it. (CC-BY is most common)
Modified	The date when the file was last edited
Keyword	The key word that will help users to find this file (e.g. Connemara)
Format	What format is the file in (e.g. CSV)
Status	This is usually set to active if the file is accessible
Update Schedule	When will the data set be updated? (e.g. As needed)

Metadata must be provided for each file that is being presented to an open data repository.

This can be supplied as a single CSV file that contains the detail necessary to describe the individual files that are to be published to an open data repository.

The file is uploaded and is then used to populate specific metadata elements for each of the published data files.

If the hosting of the files that the metadata refer to is changes you must provide a revised metadata ensure that the data remains open and accessible.

Data Categories and Sub Categories

Having a consistent category and sub category schema is critical to ensuring that data is consistent, useful and understandable. This schema should be created at the beginning of the project and constantly monitored and changed as the bulk data set grows

Over the course of the project a single Master category, Top Category + sub category per record was adopted as a standard. This was done to simplify managing the data sets in bulk form in the core data base and to also facilitate the extraction of consistent and concise sub data sets at a location and subject are level.

The bulk data base has 12 Master categories, 147 Top Categories and 1240 sub categories *(see annex 1 for the Master and Top Categories as of August 2017)*. These were not predefined. They emerged as the data was added and allocated a category and sub categories at a record level. Regular alignment and simplification exercises were carried out help maintain integrity and consistency across the master data set.

The categorisation schema used in an open data project needs to be capable of flexibly representing meaningful groups of data. This means that they must be understandable to the end user and use common terminology or if possible a standard categorisation naming convention. No such convention yet exists for rural data so inconsistency will arise between rural open data sets over time.

The categories and sub categories are used to establish and then extract subject specific data sets such as for mountains, lakes and other resources.

Other variables in the data set can also be loosely described as categories. The most important of these is PLACE as it allows all records related to a specific location such as Spiddal to be grouped for use by businesses, organisations and individuals in that specific location.

Latitude and longitude can also be used in this way to delimit data sets related to a specific area that may represent multiple locations such as the fragmented Galway Gaeltacht.

This can also be achieved in a less flexible way for more rigid areas using an additional location variable such as West Connemara, East Connemara, South Connemara or North Connemara.

There are no individual business categories or sub categories in the data sets as this would have created data protection and privacy issues best left to later in the project after the “non-Human” data sets are collected, codified and open.

Data Sources

The key lesson learnt in the collection of the data for the Connemara Open Data Project is that the more local it is the more valuable it is to local businesses and individuals. Adding very locally unique facts and information to a data element or set also increases its value to businesses.

Public Service Information is useful but rarely contributes to an individual rural businesses daily operation. It is however critical for quantifying and qualifying local broader community services and infrastructure availability.

Local Data

Most local data will be found to exist in hard copy such as books, reports, guides, newspapers, presentations. Much of this will have been produced by local creators usually as part of a one off project.

Little of the data is digitised or can be made available in machine readable form. This is usually due to a low appreciation of the value of the information being available in a digital form and accessible as an open resource. Getting permission to digitise and open the data is generally not a problem.

For topographical, historical, heritage, culture, traditions and other data sets that may be desired local knowledge and third party information sources are can be found. Again there is the challenge of normalising the data for consistency of categorisation, spelling and allocation to a location. Some data sets will have to be created from scratch through field work, surveys, observation and research.

Photos in support of a data set (e.g. Mountains, monuments, buildings, businesses etc) can be problematic. In Connemara the average age of the photo stock in use was between 5 and 8 years. Where possible new photos were taken or up to date third party ones sourced for use.

How these are processed for reuse is a choice that must be made. Most of the photos supplied in the Connemara Open Data Project are compressed 400 x 300 pixel items designed for low data overhead use online in web sites. The quality is adequate for brochure use also.

Government and Local Administration data

EU, Government and local authority data is increasingly available as open data resources. During the project it has been found that national census and data sets from the Central Statistics office require little intervention other than allocating standard categories, location, dates and latitude and longitude.

However, these data sets rarely map exactly to a given location. The use of electoral districts is a common and meaningful delimiter of a local data set but rarely encompass all of a locally recognised area or location (e.g. Moycullen has 2 major and a number of minor electoral districts).

The most logical way to deal with this is to pick the largest representative electoral district (by population size) and use this as a sampler for the whole “area”. A more comprehensive but still incomplete approach is to consolidate all electoral districts to a single data set. A super set of data was however available for “Connemara” as it is closely aligned to the Galway west constituency.

Should the data be used for cross comparative purposes (E.g. Unemployment comparison between 2 locations) it is best to use percentage not real data. This is not available for most government and local authority data sets and has to be calculated and recorded. It is best to ensure that a data set with real data (the actual figure) and another for its Percentages are created and presented.

Data from public funded organisations related to heritage, culture, history etc is also an important source of potential data. Such data can be sparse for a location because these organisations have a national rather than a local remit in gathering data and information. They also prioritise and capture the highest value rather

than the biggest number of items relevant to their brief. As such they provide a useful skeleton that can be used to add more items of local importance or note to the data set. (E.g.: It is unlikely that the topless pink mermaid in Leenane would yet be in the national heritage data set)

Geo-physical and distance data

Ordinance survey maps are useful but considerable work has to be carried out to tailor the data available by adding notes, latitude and Longitude and in many cases the historical period and data for a feature of note.

Distance data can be of value to local businesses. It can be used for journey distances and timings. Location to location data can be created using Google transit. This also provides timings per mode of transport (car, rail, walk, bus, bike etc). Such data is indicative in that it cannot ever be exact as it uses distance and an average speed and makes no allowance for the circumstances of the moment. One constraint of Google transit is that it may not calculate the distance to smaller islands where a ferry journey is necessary.

Providing distance data to major attractions and points of interest related to a location is a high value data set to have available.

Personal Contributions

Contributions from individuals can be a very valuable source of locally unique information. The use of Google forms is a useful way to collect data in a standard fashion and have it automatically recorded in an excel compatible format. The contributor can be presented with a standard form that has lists of Categories, places etc that help initially align the contributed data to the master data set.

Data Management Tools

The project has used data and graphic management tools that are free or low cost. This was done to see whether these were adequate to an open data project and how they impact cost.

These are listed below.

Tool	Purpose
Chrome Browser	Access the internet
Google search	Locate likely data sets and information
Gmail	Communicate with data set owners and contributors
Google Forms	Contributions from individuals
Google Maps	Locate items of interest and Provide latitude and Longitude
Google Street view	Provide a visual support for some data records
Google Cloud	Data storage and back ups
Google MyMaps	Creation and access to online maps from excel data sets
Dragon Naturally Speaking	Voice to text to allow read text to be digitised in machine readable form
Smallpdf.com	Convert PDFs to Excel
Survey Monkey	Run surveys to create data sets
File Maker Pro 12	Hold and manage the master database, create CSV and .XLSX files
Serif Web Plus V8	Web site creation and management
Coffee Cup	Web site editing
Smart FTP	Manage the upload of data to the servers
Mail Chimp	Communication and distribution of information to stakeholders
Microsoft Word	To run spelling and grammar checks on data sets
Microsoft Excel	Capture and manipulate data sets outside the master data base.
Microsoft Visio	Create flow diagrams
Microsoft PowerPoint	Presentation creation and graphics creation
Caesium	Compression of Photos and Graphics
Serif Photoplus X8	Manage bulk renaming and resizing of photos
Paint.Net	Photo manipulation
Adobe Bridge CC 2017	Graphics Management
Agent Ransack	File finder on PC
Flipping book	E-book creation for manuals and guides

Use of these tools or their free alternatives significantly reduced the cost of the overall project.

Licenses

A number of different licenses can be used for open data. The Connemara Open Data Project uses the CC-BY open license.

This licence allows re-distribution and re-use of a licensed work on the condition that the creator is appropriately credited. This license lets others distribute, remix, tweak, and build upon the work, even commercially, as long as they credit you for the original creation.

The license deed can be found at the following link: https://creativecommons.org/licenses/by/4.0/

Disclaimers and Safety Information

A disclaimer such as the one below may be included in the description of the data set. This is done to ensure that users are fully aware that any data or information provided may not be complete, accurate or up to date. It also disclaims any liability in relation to the consequences of using the data.

(Publishers Name) provides this information with the understanding that it is not guaranteed to be accurate, correct or complete. It does not propose to be a full authoritative list of all possible items as a data set. (Publishers Name) accepts no liability for any loss or damage suffered by those using this data or information for any purpose.

Some data providers may wish to add information as to behaviour and safety in relation to some activities or features contained in selected data sets. The Connemara Open Data Project added advice and safety information to data sets such as mountains, fishing, beaches etc in order to ensure that end users understood appropriate behaviours, equipment, possible communications restrictions and emergency contact methods

Such information will vary depending on the data set. The information should be checked against official guidance sources before being published. Examples are shown below.

Beach related information: ***Please do not swim alone or after dark. In case of an emergency please phone 112 or 999. Please take all belongings and litter home when leaving the beach. This information is provided only as advice.***

Mountain related Information: ***This mountain is remote. Appropriate equipment should be worn including good hiking boots, rain gear, food and drink, maps, compass and phone. Remember to inform others of the area being visited and return times. In case of emergency phone 112 or 999. Mobile phone service may be limited or non-existent in remote areas. This information is provided only as advice.***

Fresh water Fishing related Information: ***Fishing on any freshwater lake or river in Ireland requires a fishing license. By law any Sea Trout that are caught must be returned alive. It is encouraged to follow a Catch and Release fishing practice in Ireland. Local fishing guides and Gillies are available. Some lakes have boats for hire. Appropriate clothing and equipment should be used when fishing. Weather conditions may change rapidly. Remember to inform others of the lake being visited and your expected return time. In case of emergency phone 112 or 999. Telephone coverage may not be available in more remote areas.This information is provided only as advice.***

Data Protection and Privacy

It is important to carefully define the data sets that are to be presented for open use. These data sets must comply with local and International (Irish and EU in this case) legislation. This is especially true where data sets may identify or be used to imply or construct the identity of a person or business.

Where the data can be used in that way extreme care should be taken to be legally compliant with all relevant data protection and privacy legislation. It may also impact whether the data can be presented as “open”.

The data sets presented by the Connemara Open Data project contain no information on any individual or business in the area.

This was a strategic choice which took into account the fact that many directories and other forms of information (guides, directories, brochures.) already exist that provide this information. It also means that most of the data is supportive data that deals with fixed sets (unlikely to change in any meaningful time frame). However, it is exactly this data that is most valuable locally but poorly codified and not available digitally.

These data also have the characteristic of having no privacy aspect and little need for protection as it is not related to any person, activity or process of note in Connemara or outside it.

That being said, due diligence has been carried out to check that the data sets and approach are fully compliant or exceed the legal requirement of the Irish state and the European Union.

Use and Impact of Open Data

Rural Open data is increasingly critical to Policy makers, businesses, organisations and individuals. It has a specific impact on each.

Policy Makers and Administrations

Rural and small urban areas will be increasingly disadvantaged by not creating local open data resources to support administrators and policy makers.

Decisions on such areas are now being supported by computers using algorithms, machine learning and Artificial Intelligence (AI). These need clean consistent data in order to provide a true reflection of an area and its characteristics, dynamics and composition.

Much of the data can be fed in from Public Service Information sources but additional local data can be made available from local open data resources. This can help to ensure that better, more responsive and more relevance policy and public investment decisions are made. This reduces delays in supporting rural areas and increases their immediate benefit.

Businesses

Businesses gain a double benefit from the use of open data resources.

The first is the savings that are made in sourcing data and information for use on web sites, Apps, brochures and other marketing and communications materials. Finding and tailoring content for these various uses can account for up to 45% of the cost of a project[1]. Consistent information across many businesses also helps to build a brand image for the area, increases messaging consistency and assists in attracting new local and non-local customers.

The second benefit is related to the need for better business justifications and increasing automation of the processes involved with getting bank loans, grant applications, planning permissions, certification and other interactions with external support bodies. The availability of clean, concise open data to provide content and support information to businesses reduces the time and cost whilst contributing to more professional and comprehensive applications.

Organisations

For local organisations the availability of open data can also assist in applications for grants and support from local, national and EU sources.

For Areas, towns and villages open data can help local development groups and chambers of commerce in the creation of strategies focused on development and community support. Such strategies in themselves can be converted to open data and are a powerful tool in influencing local and national administrators, policy makers and influencers. Local data adds considerable richness to such strategies as it helps identify the unique resources, challenges and opportunities (social and economic) that may exist locally.

Individuals

For individuals the availability of open data is especially important for students and educators. The ability to support school projects and course work provides a potent way to introduce the next generation of business, organisation and political leaders to open data and its benefits.

Many individuals have specific interests in local matters and today they usually document these on hard copy. Encouraging such experts and interested individuals to capture data digitally and share it openly can assist in the preservation and generation of value from intangible and vulnerable cultural, historical and heritage elements of a local area.

[1] Connemara Programme Content creation research 2016

Annex 1: Selection of Top and Main Categories.

12 Master Level Categories

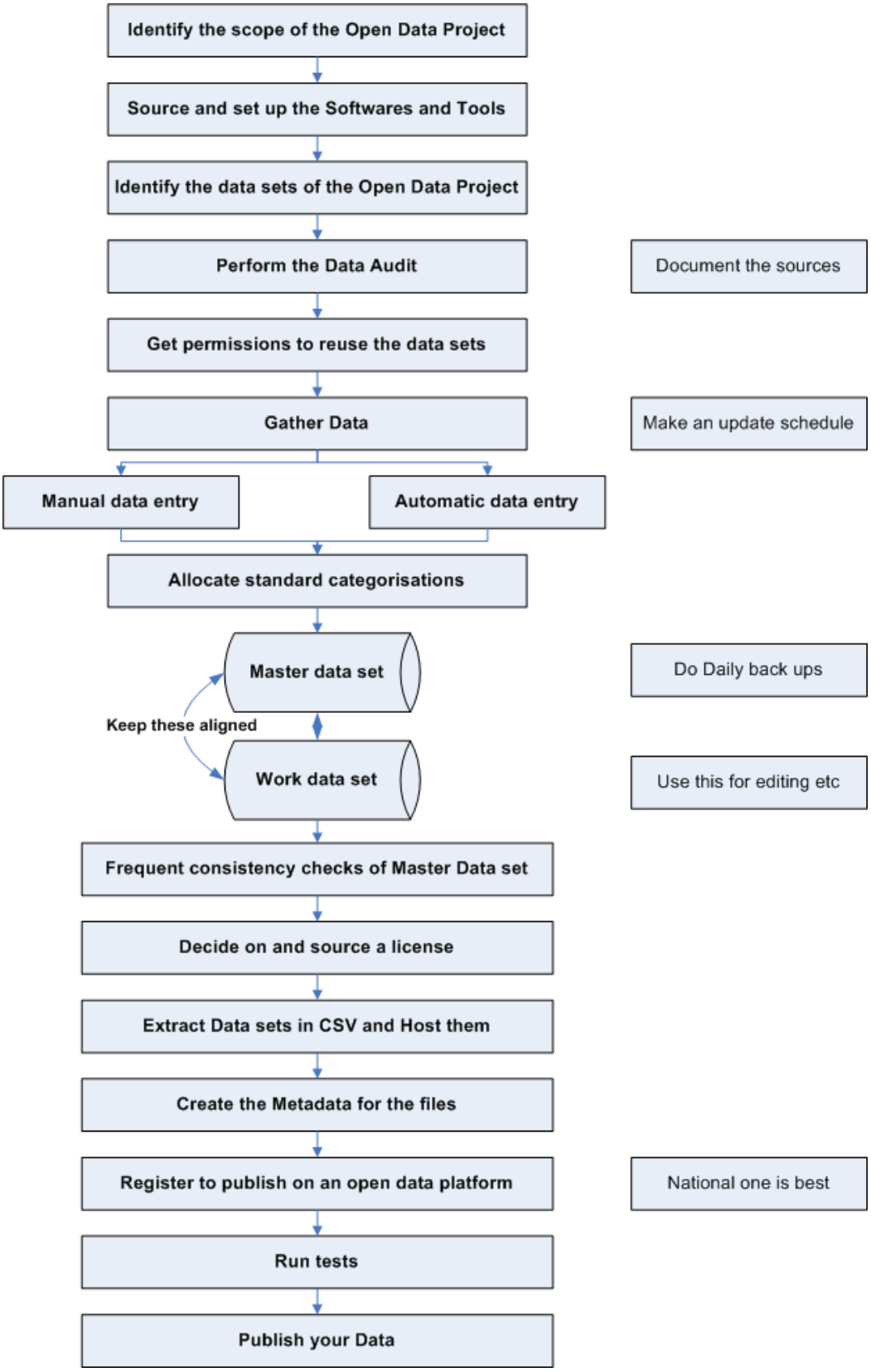
Architecture, Community, Connemara, Culture, Economy, Fishing, History, Landscape, Social, Transport, Weather, Wildlife

146 Top Level Categories

Agriculture , All Connemara %, Ambulance , Amphibian and Reptiles, Area , Artefact , Aviation, Beach, Birds, Boats , Bog, Book, Bridge, Building, Bus Stop, Business Type , Businesses , Canal , Car ownership, Cash Machine, Castle , Childrens play area, Church, Cillin, Class, Coast Guard, Coast Guard Station, Country House, Court House, Courthouse, Crannog , Crime, Cross , Dance , Defibrillator, Diet and Food, Discovery Points, Distance Kilometres, Dolphins and Porpoises, Dress, Early Christian, Economic Activity, Education, Environment, Facility, Factory, Family, Family Name, Farm Building, Film, Fire Brigade, Fish Freshwater, Fish Sea, Fishing Club, Fresh Water Fishing, Garden, Genetics, Health, Historical Event, Holy Well, Households, Housing, Internet , Internet, Island, Kiln, Lake, Lake Monsters, Land Mammals, Language Foreign, Language Irish, Library, Lighthouse, Lime Kiln, Local Festivals , Local Weather, Long Place Name, Marital Status, Mass Rock, Means of Travel, Medical Centre, Megalithic Tomb, Memorial, Midden, Military, Mill / Store, Mineral Deposit, Month Weather, Monument, Mountain, Museum, Music, National Parks, Occupation, Overview By Numbers, Parking, Personal Computers, Pier, Pilgrimage, Plants, Police, Population, Population Age, Population Birthplace, Population Ethnic, Population Nationality, Population Sex Ratio, Population Usual Residence, Post Office Services, Prehistoric Boat , Prison, Promontory Fort, Public Toilets, Railway Bridge, Railway Building, Records, Recycling, Religion, Rescue Services, River, Roadside Shrine, School, Sea Fishing Locations, Seals, Skills, Socio Economic Group, Sport, Standing Stone, Thatched Building, Time of Departure, Tourism, Tourist Information Office, Town, Town Hall, Traditions, Transit Time, Type, Unemployment, Viking, Waterfall, Whales, Work House , Wreck.

Annex 2: Simplified Process Diagram

The diagram below shows a simplified process flow of how the Connemara Open Data Project was established.



Contacts

Connemara Open Data Project

Web: www.connemaraprogramme.com/opendata

Email columjoyce@connemaraprogramme.com

Telephone +353 (0) 877 105 927

Data Sets: http://www.connemaraprogramme.com/opendata/datapage/datasets.html

Local Open Data Guide: www.connemaraprogramme.com/opendata/localopendataguide.pdf