



The Programmable City



Submission to the Department of Public Expenditure and Reform (DPER), Government Reform Unit Open Data Licences Consultation Paper

Tracey P. Lauriault, Programmable City Project
National Institute for Regional and Spatial Analysis (NIRSA)
National University of Ireland Maynooth (NUIM)
County Kildare, Republic of Ireland

Tracey.Lauriault@NUIM.ie, <http://www.nuim.ie/progcity/>

18 March 2015

1. Introduction

The Republic of Ireland, Government Reform Unit, Department of Public Expenditures and Reform (DPER) [launched](#) its first Open Data Pilot Portal data.gov.ie on July 22nd. The new portal was created by [INSIGHT Galway](#) who answered the DPER [call for tender](#) of January 24, 2014. The following suite of products were delivered on April 7th and officially approved by the Government on July 1st. This marked an important milestone for Ireland.

1.1. DPER/INSIGHT Research Documents:

1. [Best Practice Handbook](#)
2. [Data Audit Report](#)
3. [Roadmap](#)
4. [Evaluation Framework](#)
5. [Open Data Publication Handbook](#)

On the following day, July 23rd, the Government released Ireland's first Open Government Partnership (OGP) [National Action Plan](#). The newly launched data.gov.ie portal falls within Ireland's OGP Plan under section 4 entitled: Open Data and Transparency – opening-up Government data for greater accountability, improving public services and achieving economic growth. It was also something civil society, academia and government have been advocating for, for some time¹.

¹ <http://www.maynoothuniversity.ie/progcity/2014/08/sketching-the-open-data-landscape-in-ireland/>

The first draft of this plan was released at the [OGP European Regional Meetings](#) held in Dublin in May of 2014, and part of the package of deliverables listed above are related to Ireland's May 2013 commitments to join the OGP. The Government also aspires to sign onto the [G8 Open Data Charter](#). The OGP plan was informed by a number of stakeholders from civil society, academia and the public and private sectors. The documentation related to this process is available on the [OGP page for Ireland](#) and on the [Open Knowledge Foundation Ireland](#) website. Many strategies are listed in these reports, if implemented, will greatly make government more accessible and transparent to the public.

1.2. Public Service and Civil Society Engagement

Subsequently, the DPER invited members of the public service, involved with the production and dissemination of public sector data in Ireland, to attend an information session, held on the 29th of July, to discuss the documents, the portal, to solicit feedback on the development of an open data strategy for the Republic and to seek support. A public briefing was also held September 8th with civil society stakeholders and the public who have been actively engaged in the production of the OGP DRAFT Plan and engaged in Open Data and open government in Ireland.

The following are the [submissions to the consultation](#) on the Government's plans:

- [Dublinked](#)
- [Fingal County Council](#)
- [Institute of Public Health](#)
- [Local Government GIS User Group](#)
- [Dr Tracey P. Lauriault Programmable City Project NUI Maynooth](#)
- [Dr John O'Flaherty MAC Ltd](#)
- [Dr Kalpana Shankar UCD](#)
- [Heather L Wiencko](#)

1.3. Public Bodies Working Group (PBWG)

The [Roadmap](#) document proposed the creation of an Open Data Board (ODP) and a Steering and Implementation Group (SIG) to govern open data in Ireland. The ODP is currently under development and the SIG has become the Public Bodies Working Group (PBWG) on open data formed October 22, 2014. It includes public sector experts on the production, management and dissemination of data in Ireland². The Derilinx & Fujitsu Consortium was awarded the consultancy contract to coordinate the PBWG and host outreach events with the DPER. The DPER supported by

² Members include: Ordnance Survey Ireland, Central Statistics Organization (CSO), Garda, National Road Authority (NRA), Local Government Management Agency (LGMA), Wexford County Council, Roscommon County Council, Fingal County Council, Dublinked, Arts, Heritage and the Gaeltacht (AHG), National Transportation, Programmable City Project NUIM, The Digital Repository of Ireland (DRI), Environmental Protection Agency (EPA), Education and Skills, Marine Institute, Environment, Community and Local Government (ECLG), City of Cork and DPER.

Derilinx & Fujitsu has also held a number of education, training, outreach and consultation events with the private sector, civil society and public sector bodies.

The PBWG meets monthly, meetings are hosted by members at a variety of government office locations, DPER administers these, the agenda's are collectively shaped and topics are those found in the Roadman and the City of Cork offered its Alfresco infrastructure to manage records, communications and documents. The PBWG has collaborated on the development of a Data Audit Survey, how to operationalize it and the DPER is testing the process in house before deployment; the development of a metadata standard is underway; the groups is working on a list of open file formats; and tool and services managed by public bodies to advance open data are being discussed and shared. The open licencing consultation is an outcome of this collaboration.

2. DPER Open Data Licences Consultation

At the February 11, 2015 [Public Service Open Data Seminar](#) held in Dublin, Minister Howlin launched the [Open Data Licences](#) paper for public consultation.

The Paper defined open data and provided background information on the topic including its benefits, a brief on the purpose of the consultation, discussed Ireland's initiative, and described the need for an open data licence. The paper, referred to the current PSI Re-use licence, the 2013 PSI Licence, a variety of Creative Commons licences, and Public Domain Dedication Licences. Examples of UK and Canada customization of the CC licence were provided along with a list of European countries that have adopted the CC-By licence. In addition, the paper pointed to licence interoperability as an issue. The paper include 14 issues for consideration as seen in Section 4 of this submission along with responses from the Programmable City.

3. Programmable City Response

The [Programmable City](#) is a European Research Council (ERC) and Science Foundation Ireland (SFI) funded social science research project under the directorship of the primary investigator (PI) Professor Rob Kitchin. The project is based at the National Institute for Regional and Spatial Analysis (NIRSA), National University of Ireland Maynooth. Dr Tracey P. Lauriault, is a post doctoral researcher and a member of the open data PBWG.

Professor Kitchin is also the primary investigator (PI) for the All Island Research Observatory (AIRO) under the management of Justin Gleeson, the Irish Qualitative Data Archive (IQDA) managed by Dr. Aileen O'Carroll, and is co-PI on the Digital Repository of Ireland directed by Sandra Collins based at the Royal Irish Academy. The Programmable City and AIRO, under Professor Kitchin's leadership developed the Dublin Dashboard with Dr Gavin McArdle. Finally the project shares the floor in the Iontas Building with the National Centre for Geocomputation (NCG) Directed by Professor Chris Brundson.

Dr Lauriault solicited input from data experts at the AIRO, IQDA, Dublin Dashboard, NCG and the Programmable City to inform this response. These organizations rely on public data to inform research and public policy, they add value to these data by creating analytical products such as maps, charts, reports, papers and software. In addition, they collaborate with public bodies to disseminate their data, to transform data for re-use, and to make them tangible and accessible to the public. The NCG collect and manage their own data in addition to doing primary research with government data and developing open source software. The IQDA neither creates nor uses government data, but is a repository of primary data collected by social science researchers.

The following briefly discusses how each of these 4 institutions and the Dublin Dashboard work with data how intellectual property issues are addressed. This will be followed by responses to the Issues for Consideration in Section 4 of this submission.

3.1. Programmable City

Overview

The [Programmable City](#), as discussed is a publicly funded social science research project by the ERC and SFI. Data generated during the course of research by the PI, post doctoral researchers and PhD students will be archived and deposited in the IQDA.

Research Data

The research data generated by researchers are qualitative in nature and are collected via interviews with key experts from all sectors on topics related to the project. Experts sign a research ethics and consent form approved by NUIM and the ERC. Experts decide how they would like to be referenced and attributed and determine how their shared knowledge can be published. In addition, experts are afforded the right to remove themselves from the research project at any time, requiring the deletion of all files pertaining to those interviews. The data are audio recordings captured as .wav files and some are converted into .mp3 files. Interviews are anonymized, encrypted and stored on non networked hard drives and storage devices. These recordings are then transferred to a transcriber via the use of an encrypted cloud service and the transcriber has signed a non-disclosure agreement. Transcripts are then stored in the same way as the audio files. At the end of the project, all files will be deposited with the IQDA and made available to other researchers.

Intellectual Property (IP)

The data are the IP of the researcher who collected them and the PI, and as discussed can be re-used according to the wishes of the respondents (experts) as captured in the consent forms which include accepting that their knowledge is deposited into the IQDA. Once the data are deposited into the IQDA, the researcher stipulates Terms of

Use (TOU) to the IQDA in a depositors licence³ and the IQDA then makes those data accessible from its website. Researchers apply for access and must fill out an Access Request Form and accept the Agreement on Conditions of Use⁴.

3.2. Dublin Dashboard

Overview

The [Dublin Dashboard](#) is a partnership project between the Programmable City Project, the All Island Research Observatory (AIRO) and is supported by the Dublin City Council (DCC). The Dublin Dashboard is funded through an advanced investigator award to Professor Rob Kitchin from the ERC and SFI. The Dashboard was developed at the National Centre for Geocomputation (NCG) and NIRSA at Maynooth University.

Data

The Dublin Dashboard does not store any data, it either hyperlinks to existing datasets, maps, apps and visualizations; or embeds the code of the visualizations produced by the AIRO and re-disseminates these within the Dashboard's interface; or it pulls existing data from the APIs and repositories of other sites and renders them into a visualization such as a new chart or a map. In all cases, users of the Dublin Dashboard can only download the data by going to the originating source which may be a public service organization, a national trust, an international institution, a civil society organization, the private sector or a consortium comprised of public, private and academic sectors. Users can however download, depending on the module, an image of that visualization.

The data accessed from the Dublin Dashboard originate from the following organizations: Dublin City Council (DCC), Dublinlinked, Environmental Protection Agency (EPA), National Roads Authority (NRA), Central Statistics Office (CSO), Eurostat, and numerous government departments at all levels, the Ordnance Survey Ireland (OSI), Open Street Map (OSM), GoogleMaps, DublinBikes and Weatherunderground (wunderground) to name a few.

In addition, some data are collected by one institution and then aggregated and re-disseminated by another which then gets visualized in the Dublin Dashboard, for example the *Dublin Fire Brigade Ambulance Call Outs* are collected by the Department of the Environment, Community and Local Government and then visualized in the *How's Dublin Doing* module of the Dublin Dashboard. The Dublin Dashboard also hyperlinks to other data visualization services such as those provided by the Dublin Airport Authority, Irish Rail, MyHome.ie, Myplan.ie and FixYourStreet without pulling their data into the dashboard.

In other words, the Dublin Dashboard coordinates access to a variety datasets produced by myriad institutions and makes them visually accessible, readable and

³ <http://www.iqda.ie/content/deposit-data>

⁴ <http://www.iqda.ie/content/access-data>

readily discoverable via its interface. Access to the underlying data sources are provided either via hyperlinks in the interactive graphs and maps or through the *Dublin Data Store* module which hyperlinks out to the data providers used in the Dublin Dashboard.

Software

The Dublin Dashboard software was developed by Dr Gavin McArdle under the directorship of Professor Rob Kitchin in collaboration with DCC and AIRO and in consultation with a variety of end users and stakeholders. It is optimized to work with most modern web browsers (Firefox 3.5 or above, Google Chrome, Internet Explorer 9 or above, Opera 10, etc.). Some mapping modules require an up-to-date Adobe Flash Plugin and for the best results Google Chrome is the recommended web browser. Each module and sub-module may be rendering the data visible via a variety of software either hosted by its originating institution or using libraries such as OpenLayers and Highcharts to render data, accessed via API, in map and graph form respectively. The data can be combined with other data sources such as OSM and Google Maps to add spatial context. Alternatively, the data may be accessed via an API or webservice and rendered into software hosted by AIRO such as Tableau or ESRI web mapping software.

Intellectual Property (IP)

Data

As just discussed the Dublin Dashboard is coordinating access to data from a number of sources but does not host data, therefore the data IP belong to the originating source and the Dashboard adheres to their Terms of Use (TOU). Irrespective, since the Dublin Dashboard is a non commercial research project funded by public granting institutions. Therefore the Dublin Dashboard adheres to the non-commercial re-use stipulations or to the academic qualifications of most public institution TOUs.

If a user chose to download any data, they would adhere to the TOU of the source organization, for example if Open Street Map (OSM), it would be the Open Data Commons Open Database License (ODbL), if the OSI and depending on use, users would adhere to the TOU of any number of their suite of licences and depending on the data they may have to pay for these⁵, if an Irish public sector organization, for the moment that would generally be the PSI-Reuse licence, and so on. Any one module in the Dublin Dashboard may include numerous datasets from multiple sources and had it been a commercial project, access to all data would have to be negotiated with the originating institutions and some royalties might have to be paid.

Software

In terms of the dashboard itself, the following is provided on the website

© Maynooth University. All rights reserved. | Design: HTML5 UP |
Images: fotogrph | Map Icons by Freepik from www.flaticon.com is

⁵ <http://www.osi.ie/Education/Copyright.aspx>

licensed under CC BY 3.0. | Disclaimer | Contact us by email
feedback@dublindashboard.ie

The software (OpenLayers) used to render the maps in the *Dublin Real-Time* module is provided to the Dublin Dashboard under a 2-clause BSD License while the software used to render the interactive charts (HighCharts) in the *How's Dublin Doing* module is provided under a Creative Commons Attribution-NonCommercial 3.0 License.

Any software specifically developed for the Dublin Dashboard and its underlying code are currently under the exclusive copyright of Maynooth University, although another more progressive licence is under consideration. The images and icons used in the Dashboard interface are under a CC By 3.0 licence. In addition, Maynooth University absolves itself of any liability to end users should the data be inaccurate in a disclaimer⁶.

3.3. All Island Research Observatory (AIRO)

Overview

The [All-Island Research Observatory \(AIRO\)](#) undertakes academic and applied mapping research and produces spatial datasets and specialist tools to aid in their analysis and is the leading spatial analysis and planning unit within the NIRSA.

AIRO provides a suite of free public mapping and data visualisation toolkits aimed at improving evidence informed planning and also undertakes contracted applied research and consultancy projects in the area of socio-demographic and economic analysis, spatial planning and environmental analysis. AIRO also delivers a series of training and development workshops in the area of geographical information systems (GIS) and evidence informed planning.

Funding

AIRO is funded by a combination of national and international public research grants, contracted applied research by the public and private sectors and revenues generated from its Data for Decision Making seminars⁷.

Data

AIRO actively works to maximize the usage and benefit of publicly funded and readily available datasets in Ireland from private, non-profit and government sectors. An examination of its partner organizations⁸ provides insight into many of the

⁶ <http://www.dublindashboard.ie/pages/disclaimer>

⁷ <http://airo.maynoothuniversity.ie/about-airo/clients>

⁸ AIRO partners are the Higher Education Authority; Dept. of Environment, Housing and Local Development; National Institute for Regional and Spatial Analysis; International Centre for Local and Regional Development; National Centre for GeoComputation; Dept. of Environment; Dept. for Regional Development, Northern Ireland ; Dept. of Social Protection, Revenue Commissioners; Border Regional Authority; Dublin Regional Authority; Midlands Regional Authority; Dublin City Council; Dún Laoghaire Rathdown County Council; Louth County Council; Meath County Council; Kilkenny County Council; Offaly County Council; South Tipperary County Council. AIRO negotiates the procurement of data from many other institutions and creates visualizations for data journalistic

datasets from national, regional and local authorities. The data are in geomatics formats, spreadsheets and some are digitized collections of data from narrative descriptions of regions. Data are transformed into a variety of interactive visualizations where users select the data they wish to see and create a custom view of those data which they can download. In addition, AIRO provides maps and analysis for data journalists as seen in its Irish Times series. AIRO also uses and produces framework data, data for specific research projects including data associated with regional and land use plans.

AIRO also cleans data from many institutions, georeferences these and renders them into maps and makes those mapped data available in the DataStore. Datasets are in multiple formats, such as .kml, .csv, .xls, .rdf, and .shp. In addition to maps, AIRO also make data accessible via its [Data Store](#). These are value added data that have been cleaned up, georeferenced and aggregated into a variety of geographies such as small areas, electoral districts, county, and etc. These data can be freely downloaded in .csv and .shp formats from the Data Store. Special requests can be made to access national Electoral Division (ED) or Small Area (SA) framework data⁹ As AIRO works on an all-island basis many of its datasets are a combination of resources from both Republic of Ireland (RoI) and Northern Ireland (NI). For example, the outputs of the new Atlas of the Island of Ireland is a combination of statistical datasets from the [Central Statistics Office](#) and [Northern Ireland Statistics and Research Agency](#) website linked to geographical boundaries from the OSI and [Land and Property Service of Northern Ireland](#) (LPSNI).

Intellectual Property (IP)

Most of the datasets available from the AIRO site are public sector data which are covered by the PSI Re-use licence. If users download data they need to comply with the TOU of those licences. The following is the TOU text provided on the AIRO website:

By downloading this data we grant you a non-exclusive, non-transferable licence to: Use data and images produced by AIRO and contained within this site solely for non-profit personal, research or educational use. If you want to reproduce data and images produced by AIRO in any form, or use it as a basis for your own publications, other than for non-profit personal use, you must obtain a copyright licence from relevant data agencies where applicable (OSI, OSNI, CSO, NISRA etc).

As in the case with the Dublin Dashboard, AIRO includes the following disclaimer:

The responsibility for all interpretations of the data lies fully with the data user. We make no representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability or suitability of the information or related graphics contained. Any reliance you place on such information is therefore strictly at your own risk. In

purposes for the Irish Times and is subcontracted to carry out a number of data intensive research projects which it makes accessible on its website.

⁹ These data are also freely available from the CSO.

no event will we be liable for any loss or damage including without limitation, indirect or consequential loss or damage whatsoever arising from loss of data or profits arising out of, or in connection with, the use of this report.

AIRO may receive un-aggregated data from public sector bodies, these are not shared with the public in compliance with data protection rules and in formal agreement with the public bodies that have shared their data. For instance, AIRO regularly acts as an official Data Processor (through the Data Protection Act) on behalf of data organizations and aggregates raw datasets to acceptable levels of disclosure at either the SA or ED level. Some framework data such as EDs and SAs which are derived from the OSI can be viewed on the AIRO website, but these cannot be downloaded as per the TOU of the OSI licence¹⁰, Maynooth University complies with the TOI of the OSIs Research Licence. The Data from the LPSNI and NI Statistics and Research Agency fall under Crown Copyright¹¹ and the LPSNI data are subject to the Copyright Designs and Patents Act 1998 (as amended) while statistical data are subject to the [UK Open Government Licence](#). AIRO has permissions as an academic institution to use these data for non commercial purposes and to re-disseminate these on their websites. Finally, the All-Island HP Deprivation Index can be viewed but cannot be downloaded as these are produced by a private sector entity. Trutz Hasse¹² which developed this dataset with AIRO made these data available as part of a joint research project but maintain their IP, although the TOU of their licence is not available on their website. In general, this dataset is freely available (with licence) to public sector bodies and organizations involved in local and community development.

3.4. National Centre for GeoComputation (NCG)

Overview

Since its inception in 2004 with the support of SFI, the [National Centre for Geocomputation \(NCG\)](#) has become firmly established as a leading centre for research in the field of Geocomputation, applying computational methods to large spatial data sets from acquisition to analysis, modelling to visualization. The NCG is committed to:

- Extending the understanding and utilization of the capture (acquisition), analysis, modelling and visualization of spatial data from postgraduate programs to public and international project engagement
- Applying existing geotechnologies to specific problems for enhanced decision making
- Developing new innovative technologies and applications for real-time response to real world challenges

¹⁰ <http://www.osi.ie/Education/Copyright/License-more-than-small-areas-of-OSi-digital-data.aspx>

¹¹ <http://www.nisra.gov.uk/home/crowncopyright.asp.htm>

¹² <http://trutzhaase.eu/deprivation-index/overview/>

The NCG conducts research in a number of topical areas such as crime, health, disaster management, environment including research on coastlines and forests, language, transportation (for example pavement management systems and traffic modeling), retailing, historical GIS and digital humanities. NCG research includes spatial data collection, spatial data storage, spatial data visualization, and spatial data analysis.

The NCG is actively developing advanced geotechnologies / apps for use in many of these areas through current projects. It has expertise in data acquisition technologies, spatial algorithm development and spatial modelling and visualization. In many cases software, code and algorithm development is the primary objective of the research and the data are secondary to the process. Research¹³ often involves complex statistical computations which often render the data inseparable from the software/app/platform used to analyze and visualize them. For example LiDAR-data are being used in a specially developed web-services platform¹⁴ to test a pavement management system (PMS). In this case the data and the infrastructure are the key elements of the PMS, but others can use their own data in the platform.

The NCG is involved in a number of data modeling research projects that use near real-time data from public bodies, conducts mobilities research using cell tower and other location based service data and collects data using an array of data capture methodologies and technologies that include both terrestrial (LiDar) and aerial capabilities such as a Helikite and Remotely Piloted Aircraft System (RPAS) with hyperspectral camera. The NCG also produces atlases, develops mapping models and conducts image analysis using data collected from drones (i.e. RPAS), and develops 3D photorealistic models of cities, to list some of the work being done.

Funding

The NCG is funded by a mix of private and public sector consultancy or partnership research projects and also receives public research funds¹⁵. In addition some primary research and development is done under contract by industry.

Data and Software

The NCG produces their own data, and wherever possible make these available either as part of the platforms they develop, or they deposit these along with the source code used to analyze them in GitHub¹⁶, or as part of statistical algorithms they develop in the open source R system which they either deposit in the *Comprehensive R Access Network (CRAN)* or post on the *RPubs Blog*¹⁷. This is also done with the public sector data they use, for example the *Ireland Census of Population 2011: A classification of Small Areas* which is a function that “takes a 'raw' data frame from the Irish Census, and returns a data frame of derived variables, designed to be a close

¹³ <http://ncg.nuim.ie/redirect.php?action=projects>

¹⁴ <http://ncg.nuim.ie/glimpse/auth/login.php>

¹⁵ <http://ncg.nuim.ie/redirect.php?action=research>

¹⁶ <https://github.com/chrisbrunsdon>

¹⁷ <https://rpubs.com/chrisbrunsdon>

match to those used for the UK OAC classification of UK Census Output Areas”¹⁸. In this case the code and the data are what are published on the blog.

Intellectual Property (IP)

Data

If public sector data are used, the NCG adheres to the non-commercial aspects of the TOU of the PSI Re-Use licence. In some cases researchers at the NCG work with CSO data that are the result of a special tabulation ordered by one of their research partners or clients, in that case, the IP on the analysis would belong to both, but the output of that research belongs to the organization that commissioned the work and they would be obliged to follow the TOU of the PSI Re-Use licence if they re-disseminate these data, while the NCG would follow the terms of the contract.

NCG researchers have also signed a *Certification of Appointment as an Officer of the Central Statistics Office* in accordance with the Statistics Act of 1993 as per Section 21. That provides researchers with the permission to work with CSO data that have not been aggregated, or microdata¹⁹ for specifically defined research projects. This agreement is signed and all those data have specific storage and access requirements and cannot be shared as per the contract, data protection laws and the Statistics Act. Only the results of analysis of those data are shared either in papers or maps or visualizations where the data have been aggregated or input into a model and cannot be disaggregated.

When the data are from private sector organizations, the NCG adheres to the terms of their contractual obligation. NCG collected data would be published along with research papers and as discussed re-disseminated in the software or platforms they develop.

Software & Data

The NCG primarily produces open source (OS) software and conducts analysis using OS tools although some of the firmware used may not be OS. The NCG makes their code available under GPL 2 or 3 Licence and publishes that code on its own in either GitHub, CRAN or *RPubs* blog or in papers. If the code includes public sector data, it is not published in CRAN as the editors of that repository are concerned about the 3rd party re-use of the public sector data that is covered by the PSI re-use licence. The reasoning being that CRAN editors cannot guarantee that those data would not be re-used for commercial purposes. If the data and software are part of a platform commissioned by a private sector entity, then the product belongs to the commissioner while the data, even though much value has been added remains the IP of the public sector agency and re-use or sale would have to be negotiated as these data become part of a commercial product and royalties may have to be paid or special permissions sought. The ownership of the IP on the analytical side would be set out in the terms of the contract.

¹⁸ <https://rpubs.com/chrisbrunsdon/14998>

¹⁹

<http://www.cso.ie/en/aboutus/dissemination/accesstomicrodatarulespoliciesandprocedures/policyonaccesstoresearchmicrodatafiles/>

3.5. Irish Qualitative Data Archive (IQDA)

Overview

The [Irish Qualitative Data Archive \(IQDA\)](#) is a central access point for qualitative social science data generated in or about Ireland. The archive contains qualitative datasets that include interviews, pictures and other non-numerical material. It also frames the parameters and standards for archiving such data within the Irish research community. The IQDA offers training and education programs and assists researchers with the development of data management plans, data anonymization, assessing the sensitivity of their datasets and developing an archiving strategy based on that level of sensitivity.

Funding

The IQDA is funded by public research funds and generates small revenues to cover the cost of its training and education programs.

Data

The IQDA currently contains the records (i.e. research data) from 13 social science research projects and continues to grow as further datasets are deposited. As discussed the data collected during the course of the Programmable City project will be deposited in the IQDA.

The IQDA prefers that data deposited in the archive be in the following formats²⁰: .rtf, plain ASCII .txt, xml to a document type definition such as DTD or schema, .flac, .wav; TIFF (version 6) uncompressed. The IQDA will however accept html, .doc, .docx, proprietary formats such as MAXQDA, nud*IST, NVivo and ATLAS.ti formats, as well as MPEG-1, MP3, and .aiff; .jpeg, TIFF other versions, PDF/A or PDF and .RAW and may accept Photoshop .psd files under certain conditions.

Intellectual Property (IP)

Data

Data (text, audio, photos, images) in IQDA were originally collected by social scientist during the course of their qualitative research projects. Research respondents gave consent to the depositor to have their data published in the archive under any number of conditions. The data in the archive have been anonymized and depending on their sensitivity rating, are made openly accessible to the public, or available to researchers under certain conditions. The conditions of use are stipulated at the time of deposit in an [IQDA Depositor License and Legal Agreement](#). Access provisions are stated as are any special conditions. In addition the researcher deposits along with their data contextual documentation to explain the methodology of how the data were collected, caveats, disclaimers and etc. The [IQDA Guide on Producing Contextual Documents](#) provides the researcher with information on how best to do this. Data users are asked to fill out an Access Request Form²¹, and if permission is granted to access a dataset users are then asked to fill out and agree to

²⁰ <http://www.iqda.ie/content/deposit-data>

²¹ http://www.iqda.ie/sites/default/files/IQDA_Data_Access_Request_Form.pdf

the TOU in the [IQDA Data Access Form and Legal Agreement](#). The conditions of access to particular collections mirrors the conditions on use stipulated by the depositor (conditions of use may vary across collections, although IQDA encourages all depositors to consider the least restrictive option i.e. “IQDA may distribute the dataset for research and teaching purposes without asking permission for the depositor”). The *Agreement on Conditions of Use* include attribution, storage, data security, data protection, disclaimers, the deposit of derived products, withdrawal permissions and a statement attesting that copyright belongs to the depositor of the data.

4. Issues for consideration

The issues for consideration below were copied directly from the DPER [Open Data Licences Consultation paper](#) and Section 3 of this submission informs responses.

1. *What licences are public bodies currently using? (e.g. PSI, Open, etc.). Any views on the general applicability of a preferred Open Data licence to PSI?*

As seen in the discussion in section 3, university research centres produce their own data, use, add value to, combine their data with and re-use data from the public sector and NGOs in RoI, NI and international institutions. In some cases such as the NCG and AIRO they also work under contract with the private sector or under consulting contracts with the public sector or acquire data from the private sector. The many Terms of Use (TOU) from a number of licences, agreements and contracts complicates the research process, and precludes the ability to create mashups or embed data into some products. The IQDA creates its own licence and TOUs due to the sensitive nature of the social science data in its archive.

Licence interoperability is an issue, and the problem could be diminished with a universal public sector licence with the fewest restrictions on re-use as possible. As was seen in the case of the NCG some of its code could not be deposited into the CRAN repository because of the PSI Re-use licence stipulation on non-commercial reuse. All researchers cite the data they use; therefore a requirement for attribution is not a problem. Share alike is problematic as that cannot be monitored by researchers and it forces some private sector entities from embedding those data in their products, as in the case of the NCG and in some instances for AIRO since it cannot mandate its clients, even public sector clients to do so.

There are also data licences and software licences as discussed in section 3, and the Programmable City would not only support open data licences but also open source licenses as discussed in section 3.4 with regards to the work of the NCG.

2. *CCo requires waiving of certain rights (including Attribution, although it can be specifically requested, e.g., <https://open.fda.gov/terms/>). Is this acceptable to public bodies?*

Yes, as attribution and citation can be requested.

3. *If public bodies need to retain copyright of datasets, can this be done through the CC-BY licence?*

Yes.

4. *Are we clear on the copyright owner of data generated by Public Bodies in all cases? Does it always belong to the body/department which created the data? Are there always stringent contracts in place to ensure that copyright cannot rest with an external creator such as a contractor?*

Copyright ownership is not always clear.

Data do not always belong to the disseminating body.

No, there are not always stringent contracts in place, but there are some, and new contracts and agreements can negotiate access and openness.

Copyright in a research context is not always clear, however, because of the non commercial educational use of public sector data, this is not an overwhelming concern. Attribution is however part of the merit system in academic research and therefore this becomes a means to attest to IP, in some instances copyright and this becomes obvious as part of the citation practices in academic research.

In the context of the NCG and AIRO, copyright may belong to the entity that has sub-contracted them and the TOUs would be stipulated in the contracts or agreements. Researchers often do not clarify copyright on their data, however, if data are deposited into the IQDA it becomes clear in the depositors agreements.

Contracts and agreements can however set the TOUs to protect R & D but make the data output of the research process open. This can also be stipulated in procurement agreements between the public sector and the entity collecting data on its behalf.

5. *What process/governance will be in place to ensure that data being published on data.gov.ie can actually be published under an Open Data licence? e.g. to ensure the data is appropriately anonymized if necessary (and complies with data protection) or is not breaking previous copyright rules?*

The IQDA provides an excellent example of how to anonymize social science qualitative data and to protect privacy.

The CSO has data anonymization expertise and public sector institutions could develop internal protocols under the guidance of the CSO and the Data Protection Commissioner.

There are also any numbers of methods to aggregate data into a number of geographical units. NCG as well as AIRO have expertise in this area as does the OSI and the EPA. These organization can be called upon for support and to develop aggregation protocols and algorithms in accordance with data protection laws.

*Privacy by Design*²² and *Open by Default*²³ should be the norm. However some data for national security; deemed sensitive²⁴ from a spiritual, preservation, archeological or environmentally sensitive point of view should not be made openly available to the public. Explicit decision making guidelines and decision making trees should be developed and adhered to in those instances and the reasoning for not sharing some data would have to be made explicit.

6. *Can multiple licence formats be used, depending on the complexities of each dataset? For example, if a Third Party has contributed copyright material. Can an institution release some data under an Open Data licence, with non-Open Data linked via the Open Data portal, but under another appropriate licence? Or, should only datasets associated with the recommended Open Licence be included on data.gov.ie?*

Yes. On Multiple formats.

No. On the portal only pointing to open formats and data under open licences.

Since *data.gov.ie* is aggregating the metadata from multiple sources, it would be the responsibility of the originating institution to make copyright explicit and move toward openness.

The job of *data.gov.ie* is to act as a data discovery tool for the public and should not impede access to any information resources regardless of their degree of openness or data formats. It may not be possible for the originating institution to separate information products that contain data under open and non open licences, and the public is information poorer by not being able to gain new knowledge from information products created with resources from the public purse. This would be taking the openness ideology one step in the wrong direction and this approach would be analogous to censorship based on licence openness or file format openness. In the end it is about making data more accessible, and extreme openness approach would make some data less accessible.

As procurement and norms shift toward openness, this will be less of a problem, for the moment it would not be in the public interest to impede the discovery of publicly funded data on the basis of formats and licence.

²² <https://www.privacybydesign.ca/>

²³ <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex#principle-1-open-data-by-default>

²⁴ See guidelines for sensitive data - <http://geoscan.ess.nrcan.gc.ca/cgi-bin/starfinder/o?path=geoscan.fl&id=fastlink&pass=&format=FLFULL&search=R=288863>

7. Can CCo be used for all the metadata published on data.gov.ie, with a different licence being used, if necessary, for the actual dataset on the public body's website?

Yes.

Metadata and the data they describe are different. CCo on the metadata would enable the reuse of those metadata in other contexts, such as in a federated catalogue or portal as in the Case of the Digital Repository of Ireland.

8. What arrangements should be made in respect of data that a public body sells as a means of self-funding (OSi, CRO, PRA)?

The OSI provides special academic, research and non commercial access to their data upon request. In the Case of the OSI, some data are made available for viewing as seen in with AIRO, NCG, and the Dublin Dashboard.

Until which time another financial model comes into law, and sustainable and dedicated funding comes from the public sector, or differential scaled pricing comes into play whereby large, small and medium enterprise users of those datasets agree to pay a fee but also agree to have the data open to the public, the current model will have to stand.

Many of these data are essential framework data and data that can inform important national and local decisions; they should be pointed to on the portal and be made available to academic institutions and public good organizations such as NGOs or charities for non commercial reuse.

9. Are the open licences (Creative Commons, Open Data Commons) compatible with EU and national copyright and database-related legislation? For example, under Directive 96/9/EC, provision has been made for a set of sui generis arrangements whereby the creator of a database, whether a natural or legal person, can prohibit the unauthorised retrieval and/or re-use of its contents.

We do not have the legal qualifications to answer this question.

10. What are the implications for data generated for cross-border projects (e.g., Ireland-Britain, if different licensing arrangements are in place)?

As discussed in section 3.3 AIRO currently adheres to the TOUs stipulated in the licences of the Republic of Ireland and Northern Ireland as well as other international institutions. It would however be easier if there was less licence balkanization and more licence interoperability. The UK and Canadian open data licences are based on the CC and as discussed in the Licence Consultation Paper are licences worth considering as they meet national legal requirements and are internationally interoperable.

11. Are you aware of any legal impediments to using an Open Licence for specific datasets?

Yes.

See the section 3.5 about the IQDA in this submission, and also the response to Question 5 in the issues for consideration with regard to sensitive data and national security. TOUs may be more restrictive in these instances and special requests for use of the data would need to be made. For microdata, as discussed in section 3.3 and 3.4 for AIRO and the NCG, *Certification of Appointment as an Officer of the Central Statistics Office* in accordance with the Statistics Act of 1993 as per Section 21 may also be required as per the act and as per data protection laws.

Finally, please provide your views on two proposed licensing statements:

12. Content published through the national open data portal, data.gov.ie, is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence unless otherwise stated. Content on or linked through the portal which is not covered by the CC-BY licence is clearly marked with the appropriate licence or copyright statement. Such an approach allows for a tiered approach to the openness of the data. If CC-BY is not used, the public body should explain why.

Agreed as seen in responses to Question 6 of the issues for consideration.

13. Unless otherwise noted, the content, data, documentation, code, and related materials on data.gov.ie is available with a Creative Commons CCo 1.0 Universal dedication. This dedication waives all rights to the work worldwide under copyright law, including all related and neighbouring rights, to the extent allowed by law. You can copy, modify, distribute, and perform the work, even for commercial purposes, all without asking permission. This dedication implies no warranties about the work. There is no liability for any uses of the work, to the fullest extent permitted by applicable law.

14. Some data on data.gov.ie may not be covered by the CCo dedication, such as copies of copyrightable works made available to the public bodies by private entities. Such works may be made available under the provisions of extant Copyright legislation and any relevant EU Directives. Therefore, your rights to use those works may be similarly limited. Works where CCo do not apply will be clearly marked by a warning in the relevant documentation (for example: "This data is not in the public domain. Third party copy rights may apply.").

In principle this sounds reasonable when considered with item 12 above and providing that all government products irrespective of licence and format can be pointed to or discovered from the portal as part of a public interest principle.

As discussed earlier, and taking into consideration IQDA, NCG, AIRO, the Programmable City and the Dublin Dashboard, this would be reasonable.

5. Final Remarks

We are grateful to be able to submit to this consultation and thank the DPER for conducting it. Please do not hesitate to contact us with any questions.

Key words

Open data, open government, DPER, licence, DPER, AIRO, IQDA, NCG, Dublin Dashboard, etc.

Acknowledgements

The research for this submission is funded by a European Research Council Advanced Investigator award (ERC-2012-AdG-323636-SOFTCITY).

In addition Tracey P. Lauriault would like to thank Justin Gleeson from AIRO, Ruth Geraghty from the IQDA, Professors Chris Brundson and Martin Charlton from the NCG, Professor Rob Kitchin and Dr Gavin McArdle from the Programmable City Project for their contributions.