School of informatics

THE UNIVERSITY OF EDINBURGH

Population Register Consultation

25th February 2015

LABORATORY FOR FOUNDATIONS OF COMPUTER SCIENCE

SCHOOL *of* INFORMATICS
The University of Edinburgh
The Informatics Forum
10 Crichton Street
Edinburgh EH8 9AB

Telephone 0131 0131 650 5138
Fax 0131 667 7209
E-mail wwaites@tardis.ed.ac.uk

NSHCR Consultation
National Records of Scotland
1/2/9 Ladywell House
Ladywell Road
Edinburgh
EH12 7TF

Dear National Records Scotland and Scottish Government,

I have read with interest the consultation on expanding the NHS' Unique Citizen Reference Number (UCRN) into a generalised population register [1] and have some observations. Sincere apologies for not using the respone form – I'm not very good at filling out forms. I live in Scotland so this affects me. I am also a computer scientist at the University of Edinburgh and have worked in the database field on deduplication and joining of records from multiple databases as well as paying close attention to issues surrounding security and privacy for many years. To be clear, these are my own views and not necessarily those of the School of Informatics or the University of Edinburgh.

Before going into detail, the overall picture is that the population register project represents a significant departure from the way things are now. Indeed it seems at odds with the Scottish Government's own Identity Management and Privacy Principles [2] as well as the governing SNP's own position on identity registers [3]. It is worth quoting the words of the community safety minister at the time from the latter:

> *Given the current financial climate, the UK Government should have better uses for the vast sums of money being spent on this Scheme which presents an unacceptable threat to citizens' privacy and civil liberties... This money could and should be used to pay for much more worthy causes like more teachers, nurses or police officers or more schools and hospitals.*
>
> – Fergus Ewing, February 2009

Most recently in the Scottish Parliament the deputy first minister reiterated that, "the Government opposes identity cards and does not propose to introduce any new national database" [4]. This is true but deceptively phrased. At issue is not a *new* database but creating a *primary key* that can be used to join information from different databases. This new use of an existing database has wide implications and is a sufficiently large departure from current practice that this project

should not proceed without a full debate and primary legislation rather than a quiet consultation and a statutory instrument.

The consultation document, and draft regulations show significant evidence of wooly thinking, of which several examples are given below. This begs a question about the consultation process itself. There are several possible ways to account for these defects. Either the project is poorly thought out in which case it should certainly not go ahead in its current form. Alternatively the consultation document does not reflect the true nature and scope of the project, and again it should not go ahead without a full and very thorough process of publicising and debating exactly what is proposed.

The UCRN is not anonymous

The background material of the consultation on the NHS central register contains this surprising statement, "The UCRN is an anonymous number created by the NRS for each individual" (§7). This is elaborated a bit to explain that no identifying information is embedded into the number. It is a serial number. However the very purpose of the Unique Citizen Reference Number is to have a *unique* number to *refer* to a *citizen*. There may be no identifying information embedded into the number itself but it is identifying information. By definition. That is its purpose.

Consider for a moment what is meant by "name". A name is a sequence of letters or sounds, a token if you will, that is used to refer to a person, John McColl, say. It's ambiguous though because many people have the same name. So if we want *unique* names for people, maybe we can take a whole bunch of other information about that person (John McColl from Oban born in 1860) and use *that* as a name. This is an old practice in Scotland, referring to people by adding extra tags such as where they're from, or what their father's name is. John Oban is every bit as much a name for the same person as John McColl is.

Having slightly different versions of someone's name made from different bits of biographical information gets cumbersome quickly. It is also not especially efficient to query, particularly across multiple datasets. So instead we assig a number to this bunch of biographical information and use *that* as a name. This is exactly what the UCRN does.

The UCRN is a name by which the government knows you, it is not anonymous because a name cannot be anonymous, by definition.

This point may seem unimportant but it shows either sloppy writing or sloppy thinking and sloppiness is not a desireable feature of a national population register. Particularly around such central concepts such as name and identity which should be crystal clear. Saying that the UCRN is "anonymous" and "contains no identifying information" is misleading since it itself is identifying information by design.

A trivial enumeration attack

More serious, because it deals with the actual changes to the regulations rather than simply explanatory prose, the proposed amendments to Schedule 2 of the National Health Services Register

(Scotland) Regulation 2006 contains additional text in several sections that says "any other information that is equivalent to the information that has been provided by the [person, body, local authority] but does not match that information" may be provided.

It is difficult to understand what this means, but presumably the intent is to facilitate a transaction that goes,

> **Entity A:** I have John McColl's UCRN here and I think his address is 3 Argyll St., Oban. What address do you have?
>
> **NHSCR:** That's funny, I have his addressas 3 Argyll St., Glasgow.

This doesn't seem unreasonable until the behaviour of the system is looked at in the face of unexpected input. In security design it is important to consider not only how a system is expected to work in the normal course of events but how it is expected to work in the face of an attack, when people act differently than expected. So what if I don't know John Oban's address but I want to find it out? It's very simple, just make up *any* address and the query will return the correct one.

Furthermore, the *kind* of information held in the NHSCR is likely to be well known. It is some relatively small set of "basic demographic information" (§5). So for any of those attributes the above game can be played.

Anyone with access can extract all information in the NHSCR about a particular person within the proposed rules.

This is an abuse, and should not be allowed, but a good system is robust against abuse. The mechanisms for preventing such abuse are unspecified.

Identity, authentication and data

The deputy first minister said in parliament that "the measures would... ensure that public sector organisations can verify whom they are dealing with in order to deliver the right services to people" [4] and a similar purpose is explained in the consultation (§14-16) to facilitate the development of the "myaccount" single sign-on mechanism for public sector services. This is the justification for sharing data with 97[1] different organisations.

First, §8 says that there is an opt-out, that people may choose to not use on-line public services. This is good however the opt-out should also be extended to data sharing. Or rather it should be opt-in. No data should be shared – even in the case that that a government body collected it in some other way – unless explicit permission is given.

§15 is a bit confused. It says that the "Improvement Service" who operates myaccount checks the information that someone gives when registering against the NSHCR and that this is used to identify them. From that point the Improvement Service vouches for their identity. The problem with this is that at no point is *identity* verified. What is verified is the ability to produce some "basic biographical data" and thereafter to produce a password. It is not the same thing. Consider:

[1] For some reason the numbered list of organisations starts at #2.

- Someone collects readily accessible basic biographical data on someone else and uses this to register an account
- Someone uses someone else's password – whether illicitly or licitly acquired or not

What the myaccount service does do is *authentication* and we can have a greater or lesser degree of trust in the reliability of that authentication.

It is certainly less trustworthy than the two-factor authentication used by most banks and many on-line services.

Authentication is not the same as *identification*. A person may legitimately have multiple identities. For example a civil servant carrying out an action in an official capacity is using a different *identity* from when they carry out an action in a private capacity. A landlord interacting with the council as a landlord is likewise using a different identity than when they interact with the council as a resident. This is the case even though the landlord or the civil servant might *authenticate* using the myaccount service to do both of these things.

These concepts of authentication, identity and data should be thoroughly understood and the plans that the consultation is concerned with rethought in this light.

Centralisation and honeypots

In many ways the myaccount programme [5] is more revealing about the true implications of the proposed changes with the use of the UCRN. Consider the data that is collected by such an operation in the ordinary course of business.

First myaccount collects a username and password and possibly some low-value biographical data. So what? Big deal. Then they may collect more information such as council tax numbers, photographs of identification or invoices or bank statements when people "increase" their "assurance level". Finally the operator collects information about what services have been used and when. All of this information indexed by an identifier[2]

What necessarily builds up is a database of patterns of interaction with government. This is known in the industry as "pattern of life" (but of course is incomplete because only the part of life that has to do with interacting with government). On the one hand this kind of information is *extermely* good at identifying individuals – though it is not very good for authenticating them. When someone does something out of the ordinary it might be a good idea to use that as a signal that a stronger authentication mechanism is indicated, for example.

However all of this directly contradicts the Privacy Impact Assessment [6] which says, "myaccount only holds a thin mandatory set of personal information". Actually myaccount must hold quite rich metadata about an individual's interaction with government – unless no logging is performed

[2]The Privacy Impact Assessment [6] for myaccount says that "the UCRN is not used as a Persistent Pdentifier, [it] will only be used to match at a high level with service provider records." This is nonsense of a similar kind to "the UCRN is an anonymous identifier". There is – must be – a one-to-one mapping between the UCRN and the internal identifier used by myaccount so the distinction is completely vacuous.

which isn't at all plausible. All of this juicy data on everyone in Scotland in one central place presents a ripe target. Even better that this is also the single trusted entity so a perfect vector for obtaining access to every *other* government entity.

Putting all eggs in one basket is very bad security architecture. "The information architecture... with the UCRN at the centre" *does not* "[help] to protect an individual's privacy" [6].

There are plenty of other flaws with the myaccount Privacy Impact Assessment. It would benefit from the attention of an expert in the field of data and privacy. There are alternative approaches which would be much better for the security and privacy (and cost!) for all concerned. But this consultation is not about myaccount as such so suffice it to say that here too there are serious concerns.

Time is now short for the consultation deadline and this letter has only scratched the surface of the issues involved. Hopefully I have succeeded in highlighting some of them and conveying the impression that any such project for a population register requires a much more thorough vetting and public scrutiny than the present proposal has received.

Sincerely,

William Waites CSci MBCS AMFSF
Research Associate LFCS

cc: Marco Biagi MSP; Patrick Harvie MSP; Scottish Ministers

References

[1] "Consultation on proposed amendments to the National Health Service Central Register (Scotland) Regulations 2006", Scottish Government, Consultation, Dec. 2014. [Online]. Available: http://www.gov.scot/Publications/2014/12/5990 (visited on 18/02/2015).

[2] "Identity Management and Privacy Principles V2", English, Scottish Government, Tech. Rep., Oct. 2014. [Online]. Available: http://www.gov.scot/Topics/Economy/digital/digitalservices/datamanagement/IdentityandPrivacy/IdentityManagementandPrivacyPrinciplesV (visited on 19/02/2015).

[3] F. Ewing, *SNP confirm Scottish ID card opposition*, Feb. 2009. [Online]. Available: http://www.snp.org/media-centre/news/2009/feb/snp-confirm-scottish-id-card-opposition (visited on 19/02/2015).

[4] *Meeting of the Parliament – Official Report*, Holyrood, Feb. 2015. [Online]. Available: http://www.scottish.parliament.uk/parliamentarybusiness/28862.aspx?r=9784.

[5] *Myaccount Introduction*, English, Apr. 2014. [Online]. Available: http://www.gov.scot/Topics/Economy/digital/digitalservices/Sign-intoOnlineServices/myaccountBlogText (visited on 25/02/2015).

[6] *Myaccount Policy Privacy Impact Assessment*, English, Apr. 2014. [Online]. Available: http://www.gov.scot/Topics/Economy/digital/digitalservices/Sign-intoOnlineServices/myaccountPolicyPIA (visited on 25/02/2015).