odug UK Open Data User Group UK

The National Information Infrastructure (NII)

Why, What and How

January 2015

The National Information Infrastructure (NII)

"The specification and delivery of a data infrastructure of open public data, maintained to agreed standards, which describes the fundamental components of our society and can be openly accessed, used and re-used by anyone."

Foreword

In twenty first century Britain, a citizen cannot go to a single source to find their local doctors, dentists, a list of local licenced premises, a national list of addresses or even the location of a local public toilet!

Nowhere in government is there a central list that identifies all our public sector bodies. So although most government spending is now published as open data it is difficult to work out and to compare spending between different public bodies, even if they are delivering similar services.

Is this right?

Whilst the public sector is well-organised in some areas, it is not in others and, in any event, suffers from inefficiently-connected functions, particularly where the delivery of individual services crosses multiple organisations. They can be overburdened with complexity and often impeded by politics, however good underlying intentions may be.

Different public sector organisations and different public services are invariably described and classified in different ways with no coherent method available to locate them, collectively or individually. This impedes clear, open, evidence-based decision making about how services are performing or how they could be better designed. It results in duplication of effort, gaps in coverage and difficulties in delivering commercial products and services, including those delivered to and on-behalf of the public sector.

A National Information Infrastructure (NII) of open-by-default public data (with agreed exclusions, for example relating to certain personal information) would enhance efficiency, ultimately saving costs and benefiting the public.

Unfortunately, the UK has been muddling along for years without coherent data describing the fabric of our society. Additionally, as we approach the third wave of the Internet without a nationally-recognised set of “core reference data” (discussed further in this paper), the potential for the UK to advance in areas such as “Smart Cities” or the “Internet of Things” is significantly stunted.

The above factors inhibit our ability to understand how society functions and where there might be opportunities to design and deliver more cost-effective public services; an essential task if we are to reduce the national deficit and maintain living standards.

Our national infrastructure must be made open, not, as some may fear, to foster a “Big Brother” society, but to increase our ability to make better decisions, both as individuals and collectively for the benefit of society and the economy. Transparency in relation to the choices and trade-offs being made on our behalf will improve understanding and facilitate informed, constructive discussions. An open NII can also be put to good use by citizens and businesses alike. The public would be

happier that they have the ability to find out what they might need to, and be entitled to, know without undue difficulty. Businesses would be able to make better-informed decisions, benefiting the economy as a whole.

This paper is designed to provide a greater understanding why an NII is critical, what it can assist with and how it can be delivered and maintained. It will also deal with the issues that need to be addressed to make it a success.

Simply put, there is a large jigsaw puzzle to assemble, but it is completely feasible to put the data infrastructure contemplated in this paper in place. The necessary technology is relatively inexpensive and readily available, as is most of the data, which is locked away in silos in different parts of the public sector.

The proposals herein do not represent a huge IT challenge. The Internet has made the cost of disseminating data so low that it is now uneconomical to charge for it. This provides an opportunity to build an NII at minimal incremental cost.

Delivering an NII calls for: a strong political and public sector commitment, a can-do attitude, careful planning, and proper governance. However, dramatic results can be achieved with the proper will, as has been shown historically in areas of infrastructure such as transport, engineering and telecommunications. There is no room for complacency. At present there is a lack of coherent understanding in relation to the UK's national data, which many people assume to be much better-organised than it actually is.

To summarise, a properly developed NII has the potential to deliver huge benefits to the UK at relatively little cost in comparison with most "physical" government programmes. This is at a time were, more than ever, we require more efficient ways of delivering public services and need to open up opportunities for innovation and business growth to assist in economic recovery.

This paper explains the issues and offers a way forward through the creation of an NII. The public deserve full and proper access to data which is, after all, about the society we live in and funded by their taxes.

The paper is divided into three main sections: *The Why* – for everyone; *The What* – with more detail about the proposed NII architecture; and *The How* – which explains how public sector data holders can tackle the challenge of structuring their data for inclusion in the NII. The NII structure we propose builds on an underlying platform of Core Reference Data (universal data to connect between datasets), through Subject Data (identifiers applicable to multiple datasets but not all) to Thematic Data (for example, data to describe services such as Health, Transport, Education etc.).

With thanks to ODUG members for their valuable work to help bring this paper together, please contact us to comment on the National Information Infrastructure (NII). You can find ODUG at http://data.gov.uk/data-request/blogs, email us on **nii@odug.co.uk** or contact us on Twitter **@odugUK (#UKNII)**.

Heather Savory
Chair ODUG

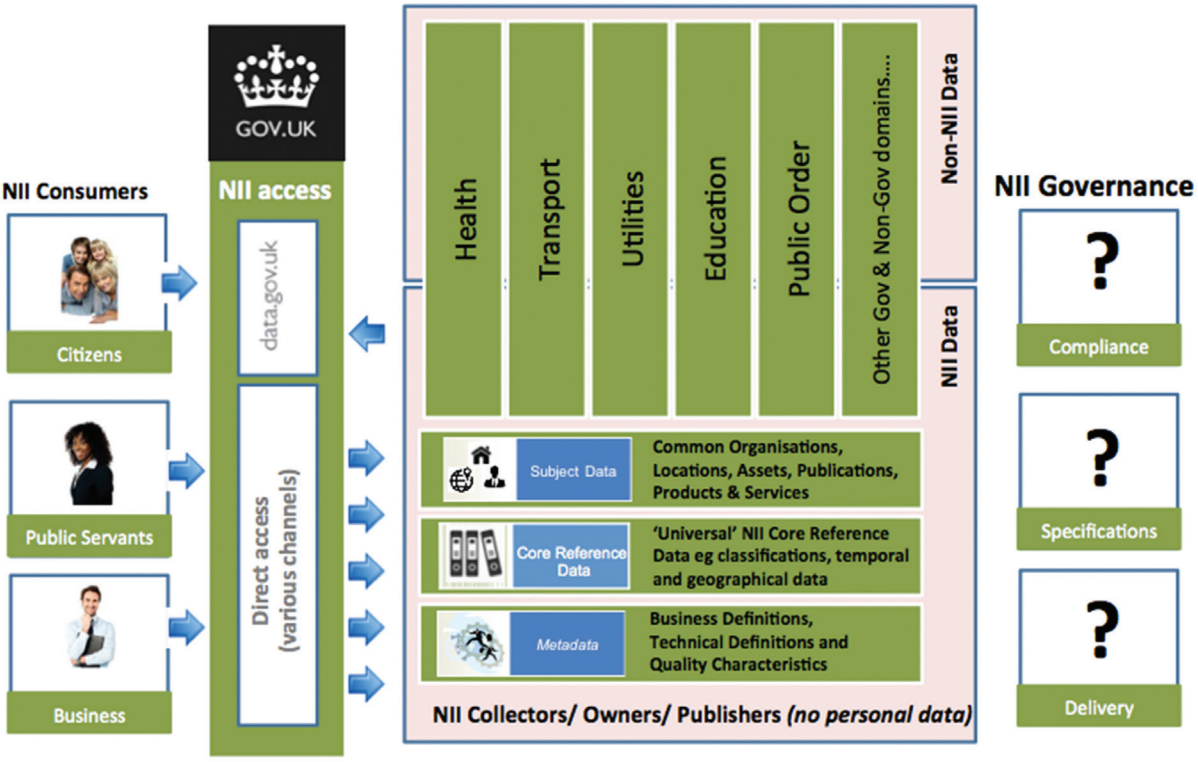
Twitter: **@SaturnSA4**

Contents

Executive Summary

As a nation, we have limited resources to deliver public services and an urgent need to drive efficiencies across the public sector. The public sector needs to innovate more to design public services which are better joined-up across functions. Service design needs to be based on well-evidenced facts and information and to use digital technologies wherever possible. This drives the need to combine management information (data) from across different public service areas in central and local government together with that from businesses delivering services on behalf of the government, including voluntary organisations, to analyse how things work now and how they can be made to work better in the future.

We are approaching the third wave of the Internet[1] without a nationally recognised set of data to describe the society we live in. Without this data, a National Information Infrastructure (NII) as depicted below, the potential for better public services or for the UK to advance in areas like Smart Cities or the Internet of Things will be fatally stunted.



The NII should be Open Data, accessible under an Open Government License (OGL)[2] and therefore is accessible by *all* for use and re-use. It is not sensible to limit the use of the NII to the public sector only, as there are significant opportunities for economic growth if we allow businesses, large and small, to benefit from the use of this data too.

1. 1=fixed, 2=mobile, 3=linked (Internet of Things)
2. http://www.nationalarchives.gov.uk/information-management/re-using-public-sector-information/re-use-and-licensing/what-ogl-covers/

Citizens should also have access to this data, so that they can hold the government to account.

The specific advantages of a properly designed NII are:

- It creates an agreed architecture for the publication of our key datasets. This framework will protect the nation's most important datasets so that they can be used by all types of organisations and by citizens
- It will make the publication of key datasets more robust, consistent and dependable
- It will be an interoperable list of individual datasets that can be linked to each other, creating new ways of analysing the work of government
- The NII will set publication standards for key public data within government, helping to 'raise the bar' for the publication of other datasets

Delivery of the NII requires:

- Overall Leadership with strong sponsorship
- Funding (at a magnitude which pales into the margins in comparison with physical infrastructure projects)
- Strategy, Planning and Execution
- A Data Governance Model
- Some good delivery exemplars to demonstrate good practice

This is not another huge government IT project. The necessary technology is relatively inexpensive and readily available, as is most of the data which is locked away in different parts of the public sector in silos. The key to the NII is recognising why open access to well maintained data and incentivising its adoption is nationally important.

A strong NII will:

- Drive the evidence-based analysis and the effective re-design of public services
- Create new economic opportunities as businesses, academics and entrepreneurs can use this guaranteed data to build new products, create new jobs and provide new products for citizens
- Enable any citizen to use the data to monitor the activities of government and hold the government to account

The UK is currently at the forefront of the international Open Data movement. A well constructed and well maintained NII will not only be the driver for the improved use of public money, economic growth in the digital product and services sector and citizen engagement; it will also create a new international benchmark for transparency.

Introduction

The Shakespeare Review of Public Sector Information[3] recommended that the government should produce and take forward a clear, predictable, accountable 'National Data Strategy'. Part of the Government's response was to publish the first version of the **National Information Infrastructure (NII)** in October 2013[4] with the stated intention that it should, 'contain the data held by the government which is likely to have the broadest and most significant economic impact if made available and accessible outside of government'.

This paper proposes moving the NII to a new level and towards maturity, supporting the Shakespeare Review Recommendations 1, 2 and 3, also explaining how the Trading Fund Data covered in Recommendation 4 is an essential component of the Core Reference Data which needs to be opened up to help release the value of other data in the NII.

However this initial, early narrow definition overlooks the **equally important benefits an NII can deliver inside the public sector**, including: opportunities for improved evidence-based policy making; better decision-making about the allocation of public sector resources to drive efficiencies in public services; and the wider benefits which can be delivered to society and citizens in their interaction with public services and their local environment.

The opportunities to drive social and economic benefit from the NII are vast; restricting its free use to government and the public sector only will limit its potential. Whilst public sector bodies can contract with external third parties to carry out data analysis we feel that the wider public good will be better served by opening up the NII to everyone. The private sector and civil society, including charities and housing associations, will find innovative ways to make use of the NII, with innovative business models. The more organisations and individuals that can have open, free access to the NII the better and the greater the benefit to all citizens.

We believe that the strategic ambition for the NII should be, **'The specification and delivery of a data infrastructure of open public data, maintained to agreed standards, which describes the fundamental components of our society and can be openly accessed, used and re-used by anyone.'**

We also stress that the NII needs to be underpinned by a set of **Open Core Reference Data**, universally adopted, thus allowing various organisations' data to be linked together for ease of interfacing, information gathering, measurement and analysis. Aside from genuine Public Records which are necessary to run society and have been in place for centuries, this data should not contain personal identifiers rather identifiers such as addresses or geospatial data (geographic co-ordinates). This is discussed in more detail later.

Public Sector Information (PSI) is collected and held as a result of the functions of government. PSI includes Public Records (for example: Births, Marriages and Deaths Registers, Land Registration Data, The Register of Companies)

3. https://www.gov.uk/government/publications/shakespeare-review-of-public-sector-information
4. https://www.gov.uk/government/publications/national-information-infrastructure

and Administrative Data held by Central Government Departments, the delivery organisations they oversee (e.g. The NHS, the DVLA) and Local Authorities so that they can run the country and carry out their various functions. The NII will primarily be composed of government data which already exists and whose compilation has been funded by the taxpayer to support government functions. The existing legal presumption[5] is that PSI should be **open-by-default** so unless the data it contains is sensitive (i.e. information about National Security or *genuinely* personal information) it should be made easily accessible and opened up for use and re-use by anyone.

The NII needs to be specified as an architecture so that it can be built properly, with priority given to the most important and useful datasets – those which describe the underlying infrastructure of the nation. This is a different approach to the initial NII which asked for all available datasets to be made open. As a result of this, we have Open Data on 2011 English Channel and Celtic Sea Plankton[6] available on data.gov.uk which may be extremely useful to oceanographers, but is not relevant to most public service delivery bodies or citizens. So the NII should not be a long list of the easy to find and publish data, it must be a collection of the most useful data and this needs to be specified. Once specified the NII must be delivered to an agreed plan, building within a framework, piece by piece, towards a mature model. To maximise the beneficial use of the NII the most important component to put in place at the first opportunity is the underpinning Core Reference Data.

Also, strong governance and mechanisms for enforcement are essential to the success of the NII, for example to agree the data content, standards and minimum acceptable levels of data quality and delivery frequency.

This paper continues with three main sections:

1. **The Why** – setting out the importance of the NII, the challenges it can help address and user examples.

2. **The What** – which details the proposed NII Architecture – the NII Data and the NII Framework.

3. **The How** – which explains the process by which data holders should determine which of their data should be included in the NII.

Section 4 covers our views on what is necessary to make the NII a reality and Section 5 concludes with a call to action to make the NII happen.

5. http://ec.europa.eu/digital-agenda/en/european-legislation-reuse-public-sector-information
6. http://data.gov.uk/dataset/2011-english-channel-and-celtic-sea-plankton-data

1. The Why

The first revision of the NII on data.gov.uk listed 233 datasets, of which 66 were unpublished. The criteria for selecting these 233 datasets were not clear and many datasets which are fundamental to the NII were not listed. Furthermore, subsequent to the initial publication of many of these datasets their maintenance has not been managed well, or at all. So the current NII Framework is not properly specified and much of what has been specified is not delivering.

The ODUG believes that the NII should be set out from first principles starting with its fundamental Core Reference Data, based on what we *should* know, to run society and make better informed decisions, rather than a reflection of the disparate data *currently* known to be available or likely to be relatively easy to publish.

Like any infrastructure, if people are to use it, the NII needs:

- Well-defined common standards
- Strong backing to make happen and to ensure its use is friction-free
- To be reliable and well maintained

Only with this level of attention will the NII reach a level of maturity, over time, which will maximise its potential benefit to society.

Why is the NII important?

The NII is important because, through transparency, it can deliver economic, social and citizen benefits. In economic terms it is a genuine 'public good', like street lighting – its value does not diminish when it is used by multiple people. The relative level of investment needed to realise the NII as a public good is, for many essential datasets, very small. The data already exists; it is just locked away in public sector silos and not made freely accessible. Thanks to available digital technologies it can be brought together as an open standard information infrastructure quite easily and made accessible to anyone over the Internet.

The NII should be a single data architecture of well maintained Open Data assets, adopted by all, which enables events and performance to be linked across service providers in domains such as Health, Transport or the Environment and in cases of, say, severe flooding all three. This concept means we can all be much smarter about:

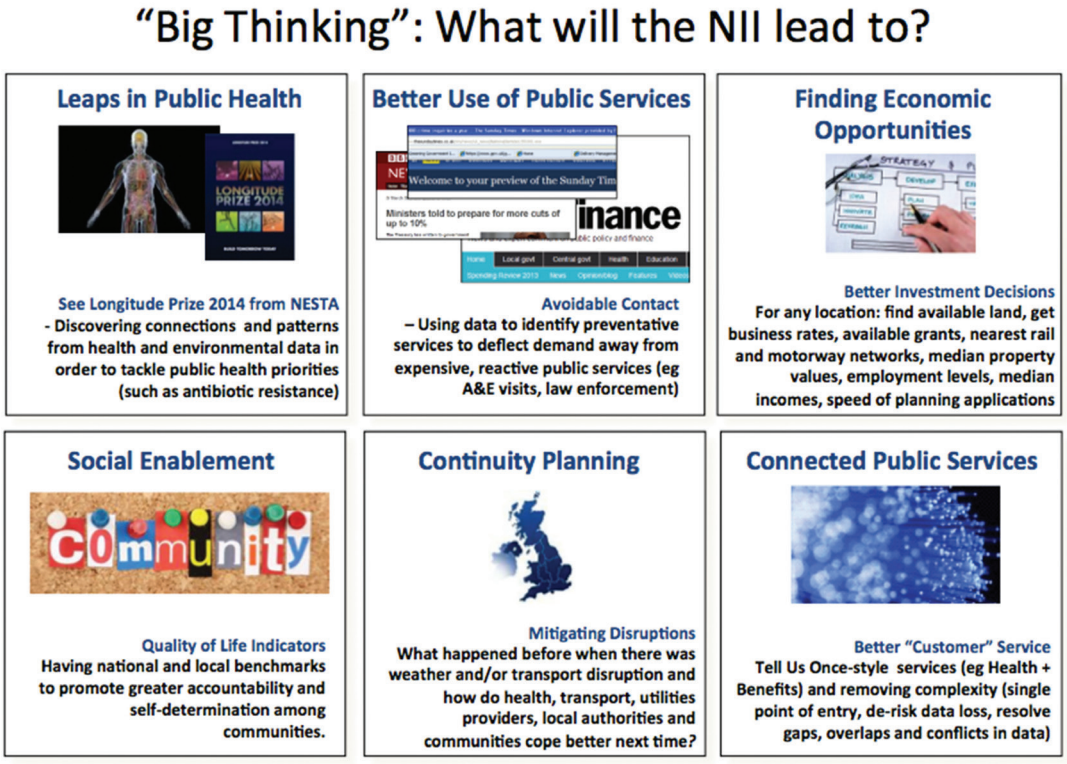
- Government transparency – allowing citizens to hold the government to account on, for example, public spending or where new physical infrastructure is planned, such as flood defenses
- Evidence-based policy making – so that public resources are invested to best effect for society
- Citizen interactions with local and national services – providing visibility on the services available and increased citizen choice

- The improved design and integration of public services – to serve the public better and drive efficiencies
- Growth and innovation in the supply of digital products and services – including those hosted on GOV.UK, Smartphone apps, improved socio-demographic research, sophisticated Big Data analysis and so forth
- Digital transformation to improve how society works, for example:
  - Smart Cities
  - The Internet of Things

The NII can deliver these wide and significant benefits by opening up the data to anyone so that it can be used to underpin decisions made about the economy, the delivery of public services, the use of public resources and where opportunities for business growth exist.

What are the possibilities?

The opportunity for the UK is significant. The examples below are just some of the key challenges which the NII can help to solve:



The potential benefits are considerable if an underpinning NII helps us make connections to ensure we are talking about the same “What”, “When” and “Where”. We anticipate the NII improving our daily quality of life in ways such as the following:

- **Leaps in Public Health**: Drug resistant diseases are responsible for 5,000 deaths/year in the UK. There have been no new classes of antibiotics developed in the last 25 years and the over-use of antibiotics globally exacerbates the problem[7]. Open anonymised health data linked to NII Data can help answers questions the researchers of these problems have, for instance:
  - What has happened, where and when?
  - How does that compare to other areas?
  - What is distinctive about those localities?
- **Better Use of Public Services**: Various estimates anticipate that the NHS will face a funding gap of up to £30bn by 2020[8]. To mitigate this, cheaper preventative service improvements could be identified through the analysis of linked Open Data if we have answers to:
  - Where services are located and what is the scale of demand there?
  - Who and where are the exemplars of Education, Health and Employment services around the country and why?
  - What is the demand on the local A&E/Care/ Welfare systems?
- **Finding Economic Opportunities**: Imagine a service that encourages investment in the right places by providing a joined up picture highlighting key facts about any locality. This needs common data used across multiple diverse domains to answer questions such as:
  - Where is the available land for development?
  - What is the proximity to transport links, skilled workforces and efficient local authorities?
  - How successful have similar businesses, especially small local businesses been there?
- **Social Enablement**: Different communities would be able to compare and contrast themselves to other areas and lobby for improved quality of life if they could determine:
  - Where local public services are located?
  - How their population/demographics compare to other communities?
  - What combinations work best and where? Can they replicate best practice examples from other communities?

7. http://www.who.int/mediacentre/factsheets/fs194/en/
8. http://www.bbc.co.uk/news/health-29735695

- **Continuity Planning:** How do we learn from major events? The direct cost of the 2013 floods was in excess of £1bn in payouts to more than 5,000 homes and a subsequent rise in their insurance premiums[9]. The indirect costs to society were significantly higher. Access to open flood risk data underpinned by detailed geospatial Core Reference Data will help to identify the actions which can be taken to protect against future flood damage if we can answer:
  - What happened, where, when and in what sequence?
  - Where were the local services and supplies?
  - Where should services be better located and configured to prevent major damage in future?
- **Connected Public Services:** As citizens we expect Government to be joined up, to know what the right response to a situation is. However without effective information flows or reliable metrics the response is often disjointed.
  - How can different organisations be sure they are talking about the same entity/place/location?
  - How can they avoid ad hoc, rushed, risky bulk data matching under crisis conditions?
  - Who owns the single record of truth?

9. http://www.telegraph.co.uk/finance/personalfinance/insurance/buildingsandcontent/10550182/Flooded-Britons-face-rising-insurance-premiums.html

Who will benefit from the NII?

The NII will benefit many different types of users, citizens, public servants, businesses and researchers in their daily lives, as illustrated below:

It is important to engage representatives of all of these groups in order to build their trust in the data and help assess demand. A deep understanding of what each group's needs are will help prioritise efforts, generate quick wins and build momentum to build the NII. A 'Build It and They Will Come' approach is likely to fail.

What do other countries do?

The UK is currently recognised as the most advanced country overall for Open Data readiness (Open Data Barometer, 2013[10]) with other countries including the US, Sweden and New Zealand close behind us. The Open Data Barometer collects information from 77 countries in total, all with a strong commitment to Open Data and moving along the same path towards a more digitally integrated society.

However the UK cannot sit still, the job isn't done. Other countries are breaking new ground to make change happen and we can learn a lot from them to see what works. For example:

10. http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf

- The Danish were particularly forward looking in 2002 when the Government provided free access to address data[11], an essential Core Reference Dataset which the UK government has yet to open up. A 2010 evaluation by the Danish Enterprise and Construction Authority[12] found that 1,200 parties were consuming the data via 22 public data distributors, with an estimated financial benefit to society of 62 million EUR, with costs to 2009 being 2 million EUR.

- Estonia is a recognised digital leader with one of the most interconnected government systems in the world. It perhaps has the most fully formed vision of an NII to date. As a general rule government systems in Estonia are not allowed to store the same information in more than one place. All government databases are linked, use common standards, and are accessible by businesses and citizens. Its central registers of core data allow challenging questions to be posed about new kinds of Government departments.

- The French government recently recognised the increasing importance of data with the appointment of a Chief Data Officer (CDO)[13] a senior position intended, 'to become involved in the policies of the production of key data in different ministries,' also, 'to contribute in the creation of true data governance, and to rationalise the circulation of data to those who need it.' The value of a senior sponsor cannot be under-estimated, but it is critical a CDO can operate as a positive catalyst for change and not become lost in the minutiae of compliance or legal issues.

How the NII as Open Data can deliver real value

Examples were provided earlier of some of the key challenges the NII can help to solve, together with an overview of additional use cases for many different types of NII user. The advantage of delivering the NII as Open Data is that this removes the friction associated with complex individual data licenses and enables different types of organisations work together to solve problems and develop value-adding products and services on top of the data.

So the NII as Open Data will allow complex cross-sector, cross central and local government policy challenges to be worked on by researchers, academics, policy officials, voluntary organisations, social enterprises and individuals, either individually or in loose, federated collaboration but where the different strands of analysis are all based on a common underlying data infrastructure. This will allow innovation, enable results to be compared, inform and improve debate about how difficult issues can be tackled and to provide opportunities to identify efficiencies in the design and delivery of public services. The NII has the potential to create common improved understanding about how society works, allow outcomes to be measured better and challenged more effectively, and provide real opportunities for social and economic benefits to arise. The NII is the management information we need to run society, and make better decisions. The wider the use of the NII the better. At present such wider use is often constrained by arguments as to who may or may not use data; who has to pay to use it, and who doesn't.

11. A 2010 evaluation found that the estimated financial benefit to Danish society of the address data alone was EUR 62m, against the EUR 2m costs incurred
12. http://www.epsiplatform.eu/content/value-danish-address-data
13. http://theodi.org/news/first-french-chief-data-officer-announced-at-odi-paris

Cross sector policy examples

As a result, at the moment, cross sector policy is dependent on individuals in departments or agencies collaborating without a common evidence base. Measurement of policy outcomes is patchy, time consuming and resource heavy. One of the key benefits the NII will offer is the ability to connect data across different government functions (provided the Core Reference Data is in place) to analyse policies and services which are delivered by multiple bodies. The development of evidenced based policy through the NII will help identify problems, focus the policy intervention and allow joint working of the multiple departments and agencies to achieve targeted, measureable and accountable delivery of policy changes.

Care Provision Example

The provision of services for an ageing population is an area the Government would benefit from being able to analyse more closely. Take care provision; a complex policy area involving Central and Local Government, the HSCIC, Care Quality Commission, Skills for Care and DCLG.

Whilst the NHS has secured initial funding for next year to sustain frontline NHS services and kick-start transformation[14] according to Sir Bruce Keogh, National Medical Director, one third of the population have at least one chronic disease which consumes 70% of the NHS budget; so resources must be made to go as far as possible. Effective policy development is essential, especially in areas like social care where the delivery may be via either Local Authority or independent providers due to differences in responsibility and accountability for the provision of care services across the landscape. Better organisation, innovation and more efficiency are desperately needed in this policy area.

Using the principles of Open Data and assuming an Open Government License for the NII Data would allow the many organisations and entrepreneurs who are working with local communities to design and provide innovative care services.

There are currently 1024 Social Care datasets released on data.gov.uk, for example:

- National Minimum Dataset – from Skills for Care
- Residential and Nursing Care – from the Care Quality Commission
- Personal Social Services – from Health and Social Care Information Centre

However without the Core Reference Data from the NII there is no way to reference and connect items of data across these datasets, making policy development by necessity opinion rather than evidence led. With the Core Reference Data from the NII it will be possible to utilise the available data to provide a granular view of the care provision for the elderly and look at opportunities which must exist to co-ordinate a change in care provision to meet user needs more efficiently, with improved evidence for the likely outcome of different options.

14. http://www.england.nhs.uk/2014/11/30/nhs-extra-funding/

As Open Data which can be manipulated and the results of analysis shared, the NII also offers opportunities for different Local Authorities to share best practice more easily.

SME Growth Example

Various government initiatives are underway to help UK SMEs and their penetration into the global marketplace. These include procurement routes, grants and advice and support to entrepreneurs starting and growing businesses.

It would be helpful for the Government to have a better understanding of the SME landscape, to assess their growth prospects and understand whether government initiatives are delivering economic value. Companies House data is already open, as is Government procurement data. These and other datasets could be linked, using Core Reference Data, to allow the analysis of SMEs by industry segment and location, which is particularly relevant where local initiatives are being supported with grants.

Entrepreneurs and SMEs without NII Data are reliant on analogue methods to identify how they should grow their businesses. The NII will allow much better analysis of UK market opportunities, providing indicators as to where to locate new businesses dependent on demographic changes. This type of analysis is largely unavailable to SMEs at present as it is restricted by data which is difficult to find, unreliable and encumbered with complex licensing issues. As there are such a wide range of SMEs their use of NII Data, underpinned with Core Reference Data to link it together, will be extremely diverse.

2. The What

What really is the NII?

We recognise how important it is to agree the terminology for the NII to help set out and clarify some concepts which are quite abstract to many people. In simple terms we need to separate the **NII Data** (datasets) a pure data infrastructure; a jigsaw puzzle of well-defined datasets from the **NII Framework** (support framework and delivery mechanisms).

NII Data

The NII should be defined as a collection of data assets, **NII Data,** which are generally created by public sector organisations in the course of delivering public services. The NII Data is specified within an **NII Framework** of tools and services to maintain the data and make it accessible

---

Together the **NII Data** and the **NII Framework** form the NII; an infrastructure, specified in a similar way to which previous generations have specified the gauge for railways, or a standard telecommunications infrastructure

---

The NII needs to be designed with care, but very little needs to be built from scratch; the building blocks for the NII are readily available as existing data which can be delivered via standard web technologies. All that is required is to establish how to link the data together in a useful way and to set expectations about the quality of the data and the frequency of its update

---

NII Framework

Looking further forward at the increasing levels of connectivity in the world, including the foundations for Smart Cities and the Internet of Things, the lack of a well-defined NII will stifle further, fundamental, digital transformation in the UK. The delivery of standard PSI and transactional services has made huge strides forward through improved citizen-facing websites for both central and local government (data.gov.uk, GOV.UK and various individual Local Authority websites). The current sites which deliver Open Data do not address the need for a coherent underlying data infrastructure. They use, generate and display data, they deliver value to citizens, but they do not approach the question as to how their data fits into the bigger picture. **There is no bigger picture defined**. The repositories are one step removed from those tasked with maintaining the data and since there is little

effective governance the value of the available Open Data tends to deteriorate over time.

The NII requires a shake-up around the perceived norms of PSI, currently collected in individual silos and in some cases added to a long list of datasets on data.gov.uk. The physical location of the data itself is not the real issue. The questions to ask (after *what* data the NII Data should include) are, for each dataset: Who collects it? Who owns it? Who delivers it? Who uses it? How is it funded? Is the cost of maintenance minimal? Then: How can it all be pulled together to deliver something really useful, the NII?

The NII Framework should support: Data Collectors, Data Owners, Data Publishers and Data Consumers

---

To make it useful and accessible the collection of **NII Data** is supported by an **NII Framework** which:

- Mandates how the individual characteristics of each dataset (metadata[15]) are presented – so that you know what is in each dataset
- Specifies the governance mechanisms to control the detailed specifications (data schemas, support etc.) and regulates where data quality is lacking or falling behind
- Sets out the delivery services which will allow data consumers to find and access the data they need (for example, search, bulk download, API access)
- Ensures that there are people with the right skills and tools available to look after NII datasets

The NII Framework should be abstracted from the underlying IT infrastructure to allow the technology-base to be upgraded over time (as better, faster, more cost effective technologies emerge).

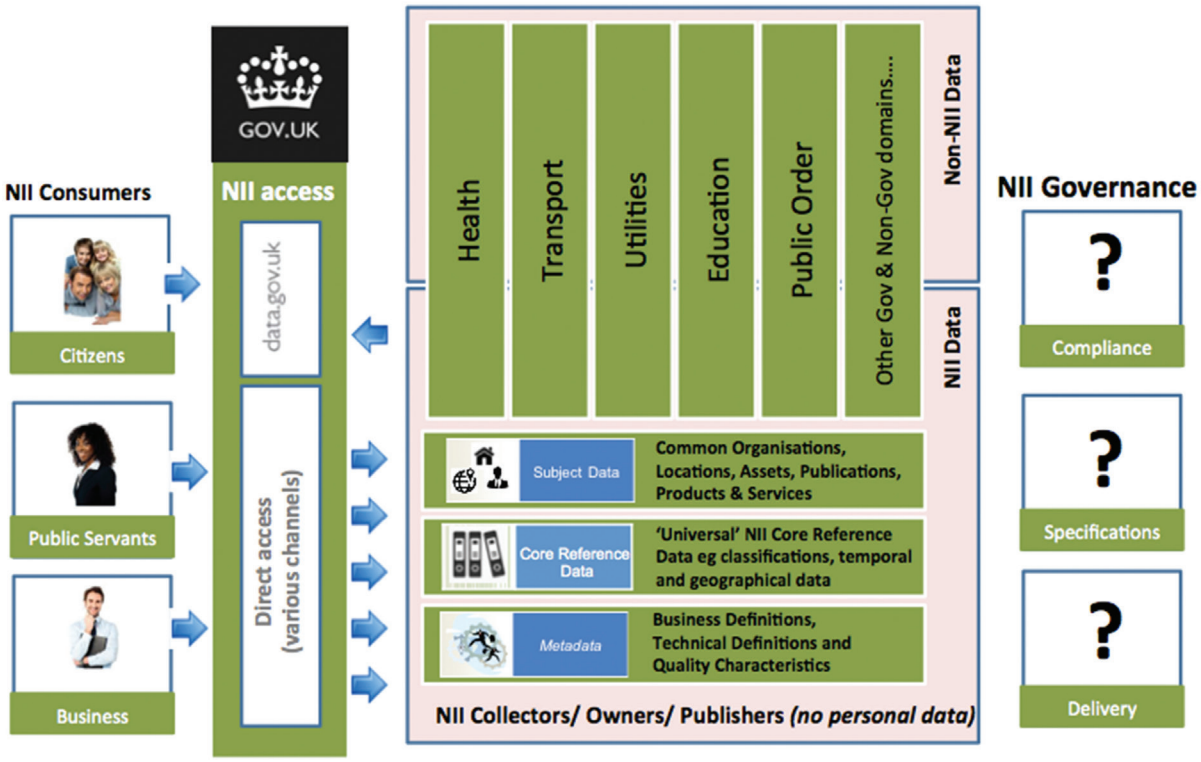
The **NII Framework** sits around the **NII Data** and specifies metadata (the meaning of each dataset); governance (over schemas, release schedules and data quality) and delivery services (how to find and gain access to reliable data)

---

15. Metadata is the data that describes a particular dataset; it provides crucial information such as who published the data, when it was published and what is the subject of the dataset. These descriptors allow people to find the data, in much the same way that an index allows you to find a book in a library. As well as descriptive data, metadata usually includes information on how to process a dataset as well, so that computers can use the data in the correct way. For example, in a column providing temperature data, the metadata will need to state whether the data is represents either Fahrenheit or Celsius. Providing this sort of descriptive data in a form that computers can read allows developers to write a programme process the data correctly and also makes it easy to link this data to other temperature datasets. Metadata is effectively the instruction manual for each dataset and a crucial element for data publication and data processing.

The **NII Data** and **NII Framework** do not specify physical system details – the underlying IT infrastructure which holds and delivers the data can be configured in many ways. Abstracting in this way allows legacy systems to be patched together and the underlying technology components to be replaced over time. Analogies with physical infrastructure would be how mobile phone numbers are transferable to new phones, or new rail carriages can run on existing rail lines – parts can be upgraded without affecting the entire system

How the NII Data and the NII Framework should be architected is as follows:



The NII Data consists only of data assets and metadata:

- The data is the middle block of the diagram above: it is arranged as a hierarchy with the most fundamental building blocks at the bottom: Metadata; Core Reference Data and Subject Data
- The NII Data consists of these horizontal platforms of underpinning data, and the vertical thematic collections of data which sit above this; for example: Health data, Transport data etc.
- The thematic collections of data will have datasets which are part of the NII Data and others datasets which are additional to the NII – the NII does not contain every single set of data for a given sector – just those which are the most important and most useful to the largest number of data consumers

NII Data example

Using an example from Smart Cities work (see PAS-182 – the Smart City Concept Model-Guide to Establishing a Model for Data Interoperability[16]) the record below gives an example of a city data entry in a dataset, in this case a report that a lamppost is not working properly.

Case No	Lamppost Ref (Object)	Location (Place)	Reported Date (Data)	Reported By (Organisation)	Fault (State)
1234	AB1234	Outside 10 The High St	10/08/2014	Epsom & Ewell Council	Light Flickering

In this example, there is data for:

- the case (EVENT) – as managed by the reporting authority
- the lamp post (OBJECT) – the uniquely identifiable asset
- the location of the lamp post (PLACE) – however based on a property description here and ideally should be a precise geo-located identifier
- the date reporting of the fault (DATE)
- the organisation reporting the fault (ORGANISATION) – again a description but should be a unique identifier
- the condition of the lamp post (STATE) – Subject Data

In the above example the OBJECT, the PLACE, the ORGANISATION and the STATE are all (or should all) be sourced from NII Data definitions so that, for example:

- All broken lamp posts can be compared to see if there is a fundamental supplier issue with the major suppliers of bulbs to Local Authorities; or
- Evidence of local crimes across the UK can be mapped to places where lighting is faulty; or
- The repair rate of this Council can be contrasted with others.

The NII Framework delivers the supporting functions:

- NII Governance – the leadership, management and governance bodies to the right in the diagram
- NII Access – the channels, access control and delivery mechanisms to the left in the diagram. We suggest that GOV.UK should act as the signposting and discovery window to the NII
- The NII needs support for structured, thematic search of the available data
- Data consumers need Delivery Services which allow bulk download of datasets, together with API support where this adds value to multiple users

16. http://www.bsigroup.com/en-GB/smart-cities/Smart-Cities-Standards-and-Publication/PAS-182-smart-cities-data-concept-model/

Census Data

We take the view that the Census Data, as currently collected, is NII Data and are in support of the government's medium term plans to use administrative data for the Census, provided that the quality and granularity of this data is not diluted or diminished as part of this process.

Local Government Data

Much of the valuable data about public bodies and public services is owned and used by Local Authorities, many of whom are now opening up more of their data than they are legally required to (by the Local Government Transparency Code[17]) and more datasets besides, because they are finding that opening their data is helpful to their work, to communication and engagement with local people and to stimulating business and enterprise.

On ODUG's recommendation this year the Release of Data Fund[18] supported a Local Government Incentive Scheme which gave small cash incentives for Local Authorities to open up some data themes in a standard way, to demonstrate that creating National Views of Local Data is relatively straightforward. These national views should also be made available as Open Data as they have value, for example for anyone who wants to work across council boundaries or to perform high granularity analysis of similar issues in different areas.

A big question for the NII Data architecture is how Local Government Data is going to fit in. This data can be considered/ classified in a number of ways:

- As a single theme – all local government data?
- As separate themes based on specific types of local environments – City/Town/ Rural?
- Separated by thematic area into the different COFOG (the UN's Classification of the Functions of Government[19]) themes

---

The NII hierarchy model needs to determine, up-front how it will include Local Government Data within NII Data

---

This is the most immediate and essential question the NII architecture needs to answer. It is not about who owns the data or where the data physically lives at the moment, but about how the jigsaw puzzle fits together.

17. https://www.gov.uk/government/publications/local-government-transparency-code-2014
18. https://www.gov.uk/government/publications/breakthrough-fund-and-release-of-data-fund
19. http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=4

What are the guiding principles?

Our view is that the NII Data should contain open national data assets to form a digital data infrastructure that ultimately includes all public records and those administrative datasets that capture information about the details of public facilities and the delivery of public services to society.

These data assets in combination can deliver transparency, social good and economic growth, form the basis for improvements in policy-making and public service design and support the digital transformation agenda, including improved citizen engagement and longer term aspirations for policy areas such as Smart Cities and the Internet of Things.

The following key principles for the NII Data and the NII Framework have emerged from a series of workshops including those recently held by the Cabinet Office.

NII Data

- Must have universally unique identifiers necessary for users to be able to reliably use, manage and monitor public infrastructure and the services they use as well as to enable interfacing to other related domains (Core Reference Data and Subject Data)

- Should be open-by-default, and only closed where there are data privacy or security constraints, however this does not exclude some closed data from being part of the NII

- Metadata must be published for each dataset with respect to data accuracy. The data quality must be maintained by the data publisher to an appropriate level. The Open Data Institute (ODI) certificates[20] work well in this regard

- The initial priority should be public data although components of the NII can just as easily include private sector data e.g. Transport Operators where they are delivering public services

- Data access should be held as close to the data source as possible since the further removed from the maintaining system it is, the less reliable the data is likely to be

Datasets should always be made available for bulk download, even in cases where APIs are developed.

NII Framework

- Must include effective overall governance and enforcement to make change happen. This is not just another IT project. The implementation of the NII should be managed and regulated by a body (or bodies) with sufficient levels of authority to ensure the integrity of the process and the deliverables

- Needs the right leadership, advocacy and evangelism from someone who ‘gets it’ perhaps at the level of a Chief Data Officer/Data Tsar with direct ministerial support as well as very senior involvement, with a formalised data role, across departments and other public bodies

20. https://certificates.theodi.org/

- There should be a defined set of specifications, including longer-term quality standards for NII Data. However although this is important for the medium term, the idea of publish first, fix later, should continue to apply to the NII for now

- The support infrastructure includes mechanisms to feedback data quality issues to the data owner, even if they are only revealed when the data is being utilised in new ways

- Should cover all thematic areas relevant to the functioning of society. The COFOG framework has been suggested covering Health, Education, Environment etc. as an independent benchmark. It is likely to need extending however; as it is recognised it does not cover all the necessary categories, notably lacking definitions for Local Government Data

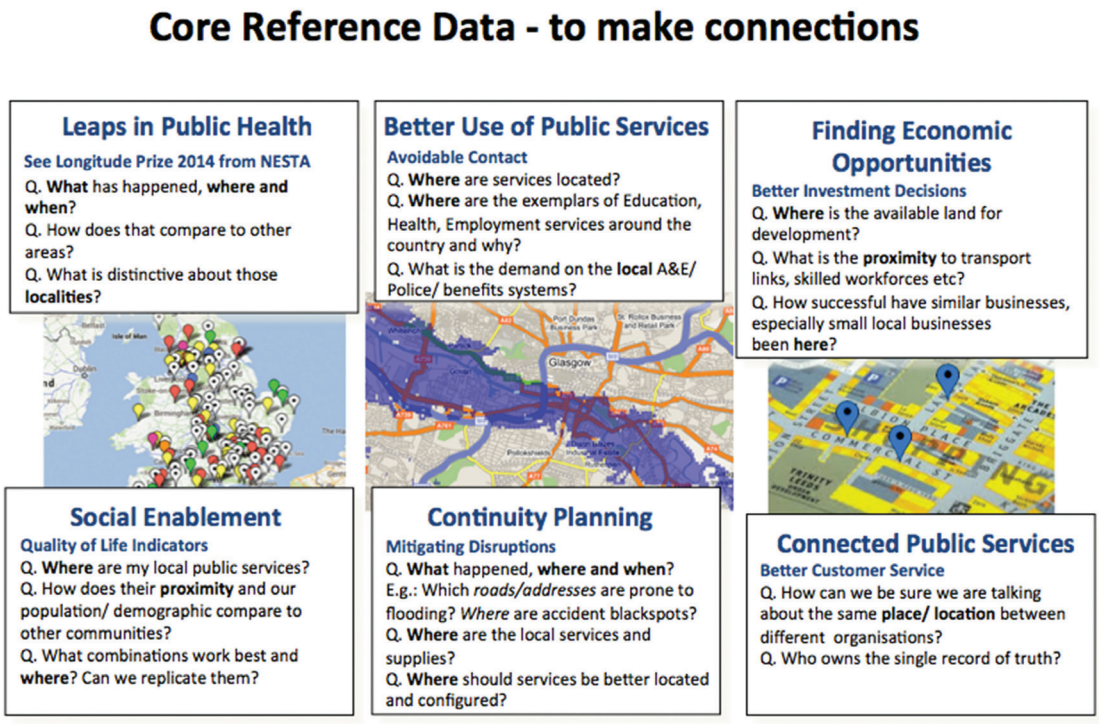
3. The How

Where to start?

The most critical NII Data is the **Core Reference Data**, **Subject Data** and their associated **metadata** i.e. the dimensions that are being managed or measured. The NII can and should extend over time as collections of NII Data are delivered and may also begin to include other types of data such as performance data (statistics) or transaction/event data, maybe even in real-time. However, any and all NII Data will be difficult to exploit without the foundations of the Core Reference Data, Subject Data and metadata that describes the contents and characteristics of the data.

Open Core Reference Data should include classifications (e.g. types of places, organisations, products, assets), Open National Address Data and Open Geospatial (mapping) Data as the essential components that underpin the entire NII. It is this underlying data that will enable linking to associate and analyse events from different but related areas like Transport, Education and Health. This is the data which will allow the analysis of multiple datasets in different 'silos' on a non-personal basis, to reveal efficiencies, inform policy-making, improve service delivery and open up opportunities for innovation and economic growth.

This diagram shows how **Core Reference Data**, in this case geospatial data, underpins the key challenges we have used earlier as examples of the benefits of the NII:



Subject Data is a less powerful set of identifiers which are common to several thematic areas and will allow the analysis of more tightly focused cross-thematic questions and areas of interest.

Examples of each type of data we would expect to see defined as Core Reference Data, Subject Data and NII Data are provided in Annex B using a real-world example of the data in the Health domain.

Any and all NII Data are difficult to exploit without the foundations of the Core Reference Data, Subject Data and their associated metadata.

Defining which datasets should be in the NII

Each thematic area of the NII needs analysis to determine how its data should be defined and structured to fit into the NII Data, which we have specified as primarily open national data assets to provide a digital infrastructure that includes all public records and captures information about the details of public facilities and the delivery of public services to society.

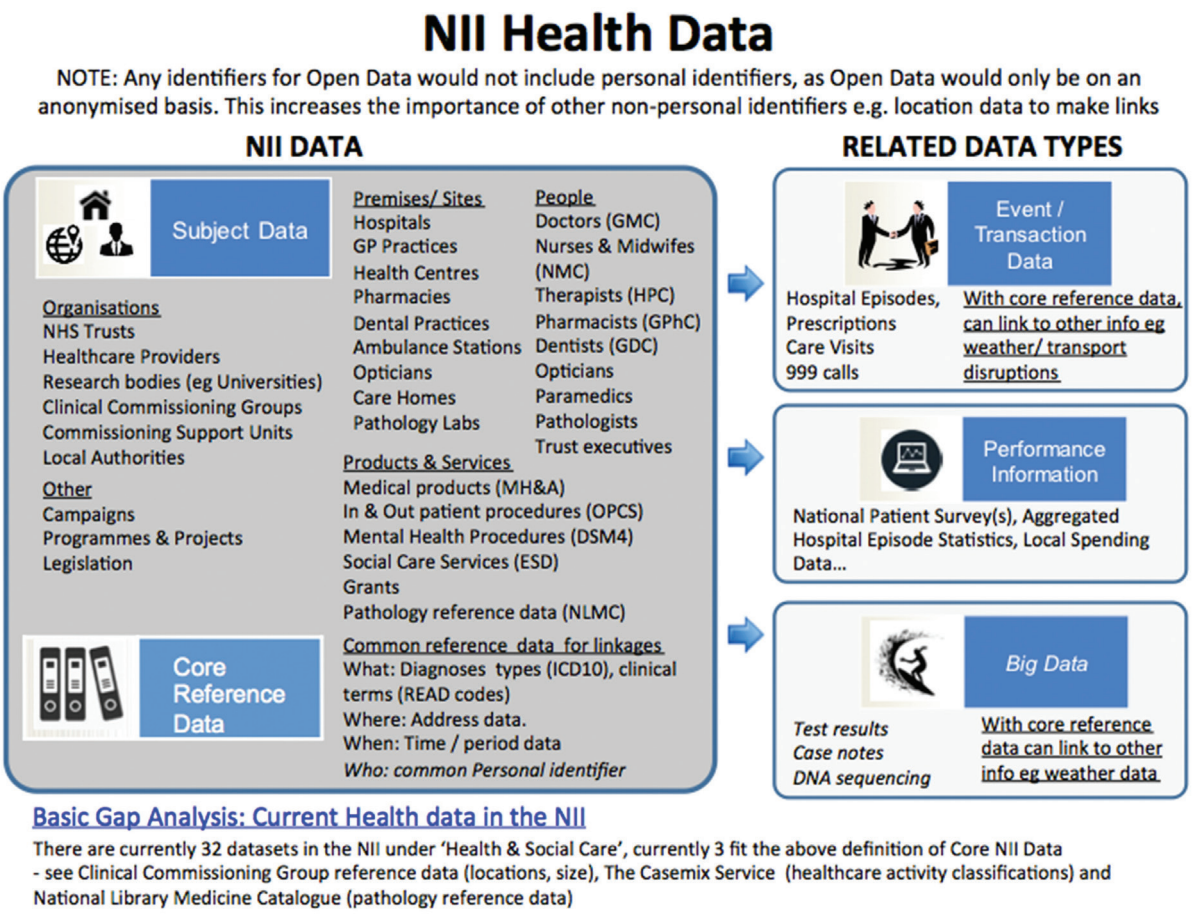
This means looking in detail at the key infrastructure and the services delivered in each thematic area, thinking about what the important components are and establishing how their data should be fitted into the NII.

The immediate priority for the NII is to define the NII Data in detail: what constitutes Core Reference Data, what constitutes Subject Data and what the NII datasets are for each thematic area.

How defining the NII Data should be approached

The diagram below is a detailed example for NII Health Data setting out what should be considered as NII Data, including individual datasets and common reference data which might later be classified as either Core Reference Data or Subject Data. For the most part the data will be PSI. Related Data such as Event/Transaction data, Performance Information and other 'Big Data' might include datasets from multiple sources which may or be PSI or other organisational or private data. Only three of the 32 datasets currently available as Health and Social Care Open Data fit the definition of NII Data outlined here.

The detailed proposal for NII Data Categories for Health Data, derived from this specification exercise is set out in Annex B: NII Data – Health Data with a Case Study on the NII Health Theme Context in Annex C. The case study is a real world example based on work being carried out at NHS England and leads to a set of working principles which show how to approach the challenge of identifying a proposed NII hierarchy for a thematic area.



The steps which need to be taken to define a thematic hierarchy are:

1. Identify available datasets, the processes generating them, and the licensing issues

2. Identify the segments of data that are likely to constitute Core Reference and Subject Data, taking the NII hierarchy model into account

3. Focus on identifying gaps and duplications – which key datasets are missing, which are too detailed to be considered NII Data?

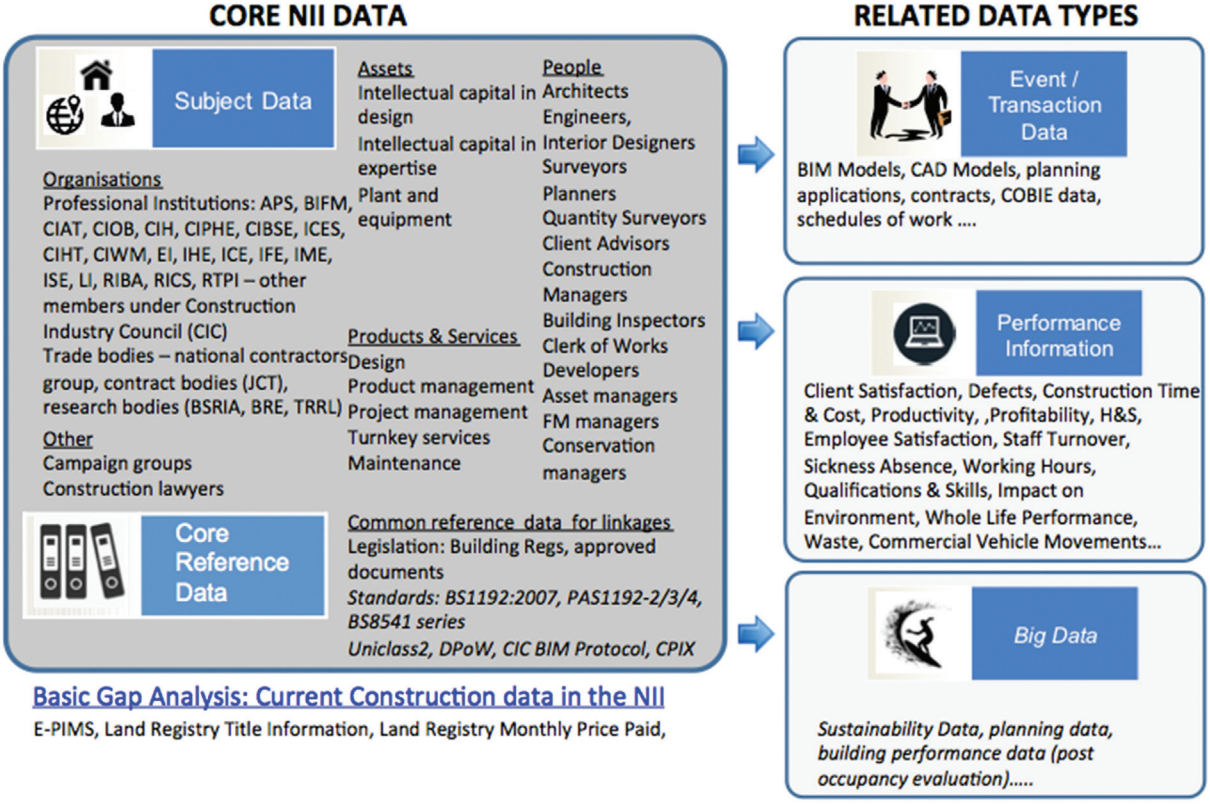
4. Agree a strategy leading to the elimination of duplications and gaps, to complete the NII Data proposal for the theme

5. Define a procedure for the inclusion, maintenance, and support of the datasets (the NII Framework)

Two other proposals are provided, the diagram below sets out the same NII model applied to Construction Data and a detailed NII Data specification for Voluntary Sector data is provided in Annex D.

NII Construction Data

NOTE: Any identifiers for Open Data would not include personal identifiers, as Open Data would only be on an anonymised basis. This increases the importance of other non-personal identifiers e.g. location data to make links



Basic Gap Analysis: Current Construction data in the NII

E-PIMS, Land Registry Title Information, Land Registry Monthly Price Paid,

4. Making the NII a reality

What is needed to make the NII a reality?

To date the NII focus has been on asking public sector bodies to volunteer their data. Doing this without an underlying set of principles and rules has not proven a great success over the last 12 months. There is no definition of how individual bodies should think about their data in the wider context, so they cannot work out what data is required, its relative level of importance, or what their delivery priorities should be. There is no overall plan for delivery priorities or timescales, collaboration between data holders or proper indication of where responsibilities lie, who has oversight in different areas and so forth.

1: Overall Leadership

The NII needs a senior sponsor who 'gets it' and has the mandate to make change happen; an owner inside government perhaps at the level of a Chief Data Officer/Data Tsar who can push it, protect it and evolve it. Initially this needs to be an evangelist, with strong governance and policy skills and decent technical understanding. Someone who will set out a strategy and a delivery plan, and engage across the public sector to make it happen.

2: Funding

No infrastructure project comes for free, there needs to be investment. However the funding requirement for the NII pales in comparison with physical infrastructure projects such as Crossrail or HS2 and, ironically, the NII would be of great of value to both these and most other government programmes. Most of the investment that the NII needs is already committed to individual data projects. Because maintaining individual un-coordinated data sets can lead to duplication, incompatibility and wasted effort a properly coordinated NII should eventually cost no more and most probably less than current expenditure on data collection and dissemination.

Funding is needed in the short term for the resource to:

- Engage the different data owners and data consumers
- Define the target sets of NII Data and implement the surrounding NII Framework
- Identify the specific priorities and engage with data owners to remove the blockers
- Push forward with both quick wins and major change

At the very least the cost which needs to be immediately recognised and funded is the cost to deliver the underlying Core Reference Data and for its ongoing maintenance.

3: Strategy, Planning and Execution

The NII architecture needs to be driven from within government against a clear understanding of:

- What the NII Data and NII Framework should comprise
- How public sector bodies will be engaged in the definition of how their data fits into the NII
- How standards will be set and who has the authority to take decisions
- What the governance and enforcement arrangements are

4: Data Governance Model

If the data is just published once it will wither and die. There needs to be at least the following in a form of an operating model:

- Data Governance Leads – representatives in each public sector data owner responsible for their data assets and committed to agreed service levels and quality levels
- A Standards Body – tasked to define what good looks like at maturity, as well as what is acceptable in the short term
- An effective Regulatory Body (which may be combined with Standards) – tasked to oversee the maintenance and availability of NII Data on behalf of all data consumers and to ensure that progress is made up a maturity curve to achieve reasonable results, in reasonable time

These stakeholders should set about overcoming the challenges of making the initial set of agreed NII definitions, namely:

NII Data

- To determine the most important datasets first: with Core Reference Data as the underlying linking identifiers for all the data which sits above it. (Data outside of the NII can still be published openly and used by those who need it, in conjunction with data inside the NII.)
- Agreeing the conceptual data architecture for each sector
- To achieve a reasonable level of consistency in the hierarchy of each sector
- To determine how Local Government data should be included in the NII

- To set agreed levels of quality and frequency of update for each dataset and ensure these are reflected in the metadata for each dataset
- To set out a path where, ideally, both bulk data and APIs are available to data users
- To assign ownership of different NII areas to responsible parties
- To assess the priorities for data publication, identifying synergies between different themes
- To put an effective governance and enforcement mechanism in place

5: Identify and deliver some exemplars

Momentum only builds when people can see tangible change. It is important to identify a set of prioritised use cases as priorities for analysis using the NII, with datasets that can be delivered as Open Data in realistic timescales. If the initial priority list is too long it risks over committing, similarly if the list is too short value will not be demonstrated.

5. Conclusion/Call to Action

We have shown here how a coherent, well maintained National Information Infrastructure (NII) can deliver significant economic and social benefits to the UK. Opening up this data in a way where is it easy and simple to use, can be passed on to others to use and can inform us all as a society will enable properly evidenced policy-making, improve transparency in public service delivery and government spending and also generate opportunities for business growth in the private sector.

We have set out that pulling this infrastructure together is well within the capability of readily available, cost-effective, mainstream technology.

We have emphasised the need for robust Open Core Reference Data (including National Addresses, Geospatial Data, Organisation and Property Registers) to allow data from different thematic areas to be linked together coherently for analysis and further analysis.

We have set out that the delivery of the NII calls for strong senior public sector leadership, funding for essential core elements, clear principles and priorities and effective governance.

Call to Action

We call on the Government to make the NII an immediate priority, and to provide a clear path forward for public sector data holders, many of whom are already moving to release more of their data as Open Data, so that their current actions will become part of a coherent bigger-picture where the parts fit together properly for the future.

We also call on the Open Data community and the wider public to press that this valuable resource be delivered for us all, as quickly as is possible.

To comment on the National Information Infrastructure (NII) please email nii@odug.co.uk or use Twitter @odugUK (#UKNII).

6. Annexes

Annex A: NII Policy Example – Housing Land Availability Assessment

It is widely acknowledged that an inadequate supply of housing land is a constraint on increasing the number of new dwellings for a growing population – a priority for the public and, therefore, all political parties in the United Kingdom. Lack of supply is a strong contributory factor to the current housing affordability crisis.

The question Government and Local Authorities must address to construct a coherent plan for the supply of new homes is: Where should new housing be built? Getting the answer to this question right is hampered by a lack of coherent information on building land availability.

This is confirmed by the number of house builders who advertise on billboards, television and through other media: 'housing land wanted'. Other organisations such as the BBC have run campaigns to crowd source information on 'wasted land' which could be used for house building or other beneficial purposes.

This raises the questions: Do we, as a nation, really not know what land is available for building more houses? Or, do we know but have a defective information infrastructure that prevents the necessary information from being brought together as an open resource? We argue that it is the second of these. All the information necessary to prioritise potential house building sites nationally exists in different places, however our current information policy prevents it from being brought together and used for this purpose.

The information sources

OS MasterMap

It may be argued that Great Britain is the best mapped country in the world at a large scale. Ordnance Survey (OS)'s MasterMap is a unique database structured as a set of closed areas (polygons) each of which is classified. Every building, every piece of road, every stretch of water is recorded in MasterMap. Therefore every potential building plot in the country exists on MasterMap, usually as a set of adjacent polygons. Because MasterMap is a database, not a 'map', it is possible to select all polygons with a particular classification. So, for example, you can retrieve from MasterMap every polygon that represents a building.

In some cases it is possible to identify what the building is, but this is not information that is stored in the MasterMap database, however every building or part of a building (where one building is represented with multiple polygons) has a unique identifier, the TOID (TOpographicIDentifier). This allows anyone, including OS, to record information against any polygon in the OS MasterMap database, not just buildings. For many buildings OS also hold an address and a UPRN (Unique Property Reference Number).

As every polygon that makes up the land surface of Great Britain is recorded on MasterMap it becomes almost trivially possible to identify land not suitable for building. For example every polygon which is classified as water, railway, or highway can be excluded from any first search for housing land, because, with the exception of large scale regeneration, the cost of building on these features would be prohibitive.

If one were to search the OS MasterMap Database for these three feature types and they were plotted as a map what would emerge would be a series of networks dividing the land surface up into closed areas. The characteristic of these closed areas is that they are not highway, railway or water.

These closed areas are known as ‘blocks’, an American term relating to the common rectangular street networks making up many American cities. In that case a block is an area of land surrounded by streets. In our case it is any area of land entirely surrounded by streets, water, or, railways as well as any other impassable feature such as a cliff.

OS MasterMap also allows us to know what is happening in the blocks. So by plotting buildings within the blocks it becomes easy to calculate how built up they are. This is done by, again, a trivially simple piece of analysis, calculating the total area of the buildings in a block and dividing it by the area of the block. This produces a statistical measure that can be expressed as a percentage of how built up a block is. While that information does not exist explicitly in MasterMap, it can be very easily inferred.

Much, much more can be inferred about the characteristics of blocks. Building size and configuration make it relatively easy to guess what a building is. Detached, semi-detached and terraced houses can be identified. Garages and outbuildings can be inferred fairly reliably. Commercial buildings such as supermarkets, warehouses or factories can be identified. By combining this information, blocks can start to be classified. Only certain types of block will be suitable for new housing. These can be filtered out to provide a short list of potential sites.

Whilst this analysis can be carried out using OS MasterMap, other than in research studies or proofs of concept, it has never, to the best of our knowledge, been done comprehensively to identify potential housing land. This is because, even though the OS MasterMap database is extremely detailed, there is information missing from it and the additional data it is held by other organisations, or part of a different data model.

Other data

The other data that could be useful for a national housing land availability model includes:

- The Department for Communities and Local Government (DCLG) data on restricted and protected land including green belt, SSSI (sites of special scientific interest) and AONB (areas of outstanding natural beauty). These could, again, be plotted on MasterMap so potential housing blocks could again be filtered to exclude those.

- The Environment Agency data on water courses and flood risk which can again be added.

- The Department for Transport data on transport land including roads, railways, airfields, ports.

So, if the above datasets are included in the NII as Open Data they can be combined to provide an assessment of where it is possible to build. The underlying Core Reference Data required to facilitate this is an Address dataset and OS Geospatial

Data. This is one of the many examples ODUG has cited as strong evidence for the government to make an Open National Address Dataset and detailed OS Geospatial Data available as Open Core Reference Data. These identifiers are the most important to allow cross-thematic datasets to be linked together and analysed.

Land ownership

Once land suitable for housing has been identified the process of finding owners, making offers and identifying future house building sites is dependent on data from the Land Registries; Her Majesty's Land Registry (HMLR), The Northern Ireland Land Register and the Registers of Scotland in the UK. It would be entirely feasible to make the UK's land registers open, as the cadastres are in many other counties.

Conclusion

If a decision was made to make the identification of developable housing land a priority a mechanism would be arrived at to make all the discovery and analysis discussed above quick, easy and inexpensive. That could be achieved by making all the necessary data available as Open Data under an unrestricted Open Government Licence (OGL) to allow agencies, individuals or entrepreneurs to carry out a national Housing Land Availability Assessment. Many other, associated and useful analyses can also be carried out.

Annex B: NII Data – Health Data

Data Type	Definition	NII Data	Health Sector as an Example
Metadata	Business Definitions, Technical Definitions and Quality Characteristics	Yes	• Business Data Dictionaries • Technical Interface Definitions • Data Quality metrics
Core Reference Data	Fundamental common reference data e.g. classification data, temporal and geographical data	Yes	• What: Diagnoses types (ICD10), Clinical terms (READ codes) • Where: Unique Address data • When: Time / period data • Who: Common Personal Identifier (not open)
Subject Data	The People, Organisations, Locations, Assets, Publications (eg legislation), Products and Services	Yes	Organisations • NHS Trusts, Healthcare Providers, Research bodies (e.g. Universities), Clinical Commissioning Groups, Commissioning Support Units, Local Authorities Premises/ Sites • Hospitals, GP Practices, Health Centres, Pharmacies, Dental Practices, Ambulance Stations, Opticians, Care Homes, Pathology Labs People • Doctors (GMC), Nurses & Midwifes (NMC), Therapists (HPC), Pharmacists (GPhC), Dentists (GDC), Opticians, Paramedics, Pathologists, Trust executives Products & Services • Medical products (MH&A), In & Out patient procedures (OPCS), Mental Health Procedures (DSM4), Social Care Services (ESD), Grants, Pathology reference data (NLMC) Other • Campaigns, Programmes & Projects, Legislation

Data Type	Definition	NII Data	Health Sector as an Example
Performance Data	Aggregated information and reports illustrating KPIs e.g. operational MI, spending/ fiscal data or national statistics	TBC	• National Patient Survey(s), • Aggregated Hospital Episode Statistics, • Local Health/Care Spend Data
'Big Data'	In this context, raw data that can be used to improve the quality of a service e.g. social media content, sensor readings, web logs	TBC	For example: • Test results • Case notes • DNA sequencing **This data can be linked to other info eg weather data using Core Reference Data**
Event/ Transaction Data	Data captured about an event or transaction e.g. hospital visit, train journey, property sale. Captures new relationships between entities	TBC	For example: • Hospital Episodes, • Prescriptions • Care Visits • 999 calls **This data can be linked to other info eg weather/transport disruptions using Core Reference Data**

Annex C: Case Study – NII Health Data Context

The definition of the NII comes with specific challenges in the health context, due to the complex nature of the health agencies and health services delivery.

Why the NII matters for Health

The best example of why such an approach is needed is represented by the multiple datasets used for the identification of GPs.

The implementation of the health component of the NII matters for several reasons:

- the efficiency gained by the elimination of duplication will generate savings in both financial and operational terms
- the ability to fix gaps and errors once for all will be made more effective
- communications between all levels of care, and especially between primary care operators like GPs and secondary care operators like hospital personnel, will be streamlined and less equivocal
- with its public, open parts, the NII will enable private services to be built with a trusted, clear set of information upon which to work
- this may result in increased power of choice for the patient, and might trigger the improvement of public services by means of positive pressure

Current situation

Several agencies provide the delivery of health services in England. The Department of Health (DH) is responsible for policy on health and social care, overseeing the English arm of the National Health Service (NHS) and executing its tasks using Arms Length Bodies (ALBs) which include NHS England, the Health and Social Care Information Centre (HSCIC), Public Health England (PHE), and the Care Quality Commission (CQC) amongst others.

All these agencies use and collect data to achieve their statutory and operational goals, and in some cases enrich that data. For example, CQC uses a list of GP Practices for its inspections. At the moment, there is no coordination among these bodies as far as their data collection and sharing is concerned. We see two major consequences due to this phenomenon:

- the duplication of data collection efforts
- the presence of gaps and mismatches between different datasets

Ongoing work on the NII for Health

A working group has been established at DH with the goal of defining the elements of the NII in a health and social care context. The definition of the elements of the NII is not limited to a list of datasets, but should include for each dataset:

- the processes resulting in the collection of the dataset
- the procedures required for its upgrade
- an indication of the data owner
- a list of dependencies, for example *foreign keys* or references to other datasets
- a clear statement on licensing

The working group will focus on a number of themes in a health context, derived from the COFOG schema. The list is as follows (preliminary draft):

Core Reference Data

- **What:** diagnoses types (ICD10), clinical terms (READ)
- **Where**: location data
- **When**: time/period data
- **Who**: common personal identifier (not open)

Subject Data

- **Organisations**
  - NHS trusts
  - Healthcare Providers
  - Research bodies
  - CCGs
  - LAs

- **Premises/Sites**
  - Hospitals
  - GP Practices
  - Health Centres Pharmacies
  - Dental Practices

  - Ambulance Stations
  - Opticians (?)
  - Care Homes
  - Pathology Labs

- **People**
  - Doctors (GMC)
  - Nurses & Midwives (NMC)
  - Therapists (HPC)
  - Pharmacists (GPhC)
  - Dentists (GDC)
  - Opticians
  - Paramedics
  - Pathologists
  - Trust Executives

- **Products & Services**
  - Medical products (MH&A)
  - In & Outpatient Procedures (OPCS)
  - Mental Health Procedures (DSM4)
  - Social Care Services (ESD)
  - Grants
  - Pathology Reference Data (NLMC)

In a health context the division between what is Core Reference Data and what is Subject Data might be not as clear-cut as in other contexts. There are many grey areas. The aim, however, is to make sure that there is at least a common layer that will allow health professionals to *speak* about their thematic data: vocabularies and descriptive metadata, and protocols to describe time and location, together with information standards.

The list above should not be considered a list of *datasets*, as it is more accurately a list of *themes*. Such themes could be constituted in one or more datasets, or there

could be more themes stored in a single dataset. The working group is running a data inventory with the goal of determining what data is where at the moment.

Ideal outcomes of this process

The working group is following an iterative approach:

1. The first aim is to identify the datasets, the processes generating them, and the licensing issues
2. The work will proceed with the identification of the segments of data that are likely to constitute Core Reference and Subject Data, taking the overall levels of hierarchy in the NII into account
3. The group will then focus on identifying gaps and duplications
4. As a final step, the group aims to agree upon a strategy leading to the elimination of duplications and gaps, and define a single procedure for inclusion, maintenance, and support of datasets in the health chapter of the NII.

Annex D: NII Data – Voluntary Sector

Data Type	Definition	NII Data?	Voluntary Sector Data
Metadata	Business Definitions, Technical Definitions and Quality Characteristics	Yes	• Organisation classification schema • Financial transaction classification
Core Reference Data	Fundamental common reference data e.g. classification data, temporal and geographical data	Yes	• Where: Unique Address data • When: Time / period data • Government bodies (local and central)
Subject Data	The People, Organisations, Locations, Assets, Publications (eg legislation), Products and Services	Yes	Organisations • Registered charities in England and Wales, registered charities in Scotland, Register charities in Northern Ireland, register of Community Interest Companies, register of Community Amateur Sports Clubs, register of political parties, register of royal charters, register of mergers
Performance Data	Aggregated information/ reports illustrating KPIs eg operational MI, spending/fiscal data, National Statistics	TBC	• Survey data on volunteering levels • Survey data on levels of charitable giving • Survey data on levels of civic engagement
'Big Data'	In this context, raw data that can be used to improve the quality of a service eg social media content, sensor readings, web logs	TBC	For example: • Accounting information gathered by Charity Commission **This data can be linked to other info using Core Reference Data**
Event/ Transaction Data	Data captured about an event or transaction e.g. hospital visit, train journey, property sale. Captures new relationships between entities	TBC	For example: • Contract and grant transactions with government • IATI data on international aid flows **This data can be linked to other info using Core Reference Data**

odug UK

Open Data
User Group UK