Document filename:	**Full Business Justification_SCCI0090 Amd 24-2015 v 2.0**		
Project / Programme	**Systems and Services Delivery (Operations & Assurance Services)**	Project	**Health and Social Care Organisation Reference Data**
Document Reference	**SCCI0090 Amd 24/2015**		
Project Manager	**Mike Presence**	Status	**Approved**
Owner	**Sue Millard**	Version	**2.0**
Author	**Mike Presence**	Version issue date	**11/08/2016**

Health and Social Care Organisation Reference Data (SCCI0090):

Full Business Justification

Document management

Revision History

Version	Date	Summary of Changes
0.1	15/06/2016	First draft for Full Stage submission
1.0	28/06/2016	Approved for Full Stage submission
2.0	11/08/2016	Approved

Reviewers

This document must be reviewed by the following people:

Reviewer name	Title / Responsibility	Date	Version
Danny Ruttle	Senior Technical Specialist, ODS	11/08/2016	2.0

Approved by

This document must be approved by the following people:

Name	Signature	Title	Date	Version
Sue Millard	Sue Millard	Head of Service, ODS	11/08/2016	2.0
Sean Walsh	Sean Walsh	Director, Operations and Assurance Services	11/08/2016	2.0

Glossary of Terms

Term / Abbreviation	What it stands for
Assigning Authority	The organisation responsible for maintaining a range of identifiers.
Cardinality	The number of occurrences of a given object or data item at each end of an association or relationship
Component	In XML terms this is a structural element of the document hierarchy.
CSV	Comma Separated Values. A common, relatively simple file format for storing tabular data in plain-text form, particularly well-suited to fixed length records.
Data Item Catalogue	A detailed technical description of all data items contained within the Object model which supports Organisation Reference Data.
Entity	An organisation, site or location which physically exists in the real world.
HSCIC	Health and Social Care Information Centre - a data, information and technology resource for the health and care system which plays a fundamental role in driving better care, better services and better outcomes for patients in England. Operates under the trading name of NHS Digital

HSCOrg	The name for the record class in the data model of a Health and Social Care Organisation.
HSCSite	The name for the record class in the data model of a Health and Social Care Organisation Site.
Intelligence	The ability for a consumer to interpret the Primary Role of a given organisation from the format of the organisation code.
Issuing Authority	The organisation responsible for publishing Health and Social Care Organisation Reference Data – currently the Organisation Data Service (ODS).
Meta Data	'Data about data' - in this context additional information about an organisation or site entity (i.e. indicating its Roles, Relationships, etc).
Multiple-frame Identifier Structure	(See also 'Unified Identifier Structure') Historically, multiple structures have been used for Organisation Reference Data identifiers to denote organisation types (referred to as 'code-frames'). For example NHS Trust codes are 3 characters long and begin with an 'R'. Code frames also convey organisation-to-site relationships
NHS Digital	The trading name for HSCIC
ODS	Organisation Data Service[1] – part of NHS Digital responsible for the publication of organisation and practitioner codes, and for the national policy and standards with regard to the majority of organisation codes. These code standards form part of the NHS data standards. This service was previously known as National Administrative Codes Service (NACS).
Partial	Partial in the context of this document is used to describe releases restricted to changed records, designed to allow updates only to be applied to reference data, avoiding the need to refresh the entire data set.
Record	The representation of an entity within Organisation Reference Data.
Reference Data	Data that defines the set of permissible values to be used by other data fields.
Referential Integrity	Referential integrity is a database concept that ensures that relationships between tables remain consistent. When one table has a foreign key[2] to another table, the concept of referential integrity states that you may not add a record to the table that contains the foreign key unless there is a corresponding record in the linked table.
SCCI	Standardisation Committee for Care Information. http://digital.nhs.uk/isce
Subject	The current record within Organisation Reference Data.
Target	Identifies the records to which relationships on the current record point.
Unified Identifier Structure	(See also 'Multiple-frame Identifier Structure') Unique identifiers, with a single, unified standard structure of 5 alternating alphabetic and numeric characters (ANANA) *with no inherent meaning*.
UPRN	Unique Property Reference Number - a unique identifier for spatial addresses in Great Britain provided in AddressBase (an Ordnance

[1] http://systems.digital.nhs.uk/data/ods

[2] http://databases.about.com/cs/specificproducts/g/foreignkey.htm

	Survey product).
URI	Uniform Resource Identifier (URI). A string of characters used to identify the name of a resource. Such identification enables interaction with representations of the resource over a network, typically the World Wide Web. The most common form of URI is the Uniform Resource Locator (URL), frequently referred to informally as a web address.
URL	Uniform Resource Locator - a reference (an address) to a resource on the Internet. For example a URL could be the name of a file on the World Wide Web because most URLs refer to a file on some machine on the network such as an XML Schema. However, URLs also can point to other resources on the network, such as database queries and command output.
XML	Extensible Markup Language. A set of rules for encoding documents in machine-readable form.

Document Control:

The controlled copy of this document is maintained in the NHS Digital corporate network. Any copies of this document held outside of that area, in whatever format (e.g. paper, email attachment), are considered to have passed out of control and should be checked for currency and validity.

Contents

1 About this Document

1.1 Scope of Document

This paper details the business justification for the changes that are being made to the interface for Health and Social Care Organisation Reference Data. It outlines the existing approach and problems, summarises the changes and the benefits these will bring, and then examines impacts, costs and funding.

Other aspects of the standard are supplied in separate documents which are detailed below.

1.2 Associated Documents

Reference	Content
Management Summary_ SCCI0090 Amd 24-2015 v 1.0.docx	High-level overview of the change from the legacy state to the new, for existing users that must transition
Technical Specification_SCCI0090 Amd 24-2015 v 4.0.docx	Technical information required to understand XML Organisation data products. An integral part of the information standard but also a live document supporting data releases
Implementation Plan & Guidance_SCCI0090 Amd 24-2015 v 2.0.docx	Guidance on implementation of Organisation data, for both new and existing users
Requirement Specification_SCCI0090 Amd 24-2015 v 4.0.docx	Fundamental Standard for Health and Social Care Organisation Reference Data (for new users)
Change Specification_SCCI0090 Amd 24-2015 v 4.0.docx	Describes the changes made to legacy products to establish the Fundamental Standard SCCI0090 (for existing users of data prior to February 2017 only)
Archive Consultation Papers[3]	Preliminary consultation papers which established known issues with legacy products and the changes needed to address them (March-May 2014)
Interface Changes Web Pages[4]	Web resources supporting transition from legacy products to the Fundamental Standard, published on ODS pages of the NHS Digital website

1.3 Document Conventions

Throughout this document, the phrase ‘Health and Social Care Organisation Reference Data’ or ‘Organisation Reference Data’ is used to refer to all in-scope data, including both Organisation and Organisation Site data.

[3] http://systems.digital.nhs.uk/data/ods/interfacechanges/consultation

[4] http://systems.digital.nhs.uk/data/ods/interfacechanges

Note that the name for SCCI0090 was previously ‘Organisation Data Service’ however with its establishment as a fundamental information standard this has been amended to ‘Health and Social Care Organisation Reference Data’.

2 Introduction

Comprehensive reference data about organisations has been maintained by a central service and provided to the NHS for the last twenty years. Over time, aspects of the data have become outdated resulting in several issues, including: the imminent exhaustion of some code ranges; inflexibility around relationships and meta-data; and a convoluted release structure.

Following a consultation period that ran between March and May of 2014, the issuing authority laid out a package of changes to the existing data interface which aim to address these issues.

These changes have been brought before the Standardisation Committee for Care Information (SCCI) in an information standard. This justification is a component of that standard and provides an outline of why the changes are required as well as a consideration of the impact and cost implications.

3 Case for Change

The approach laid out in this standard was thoroughly examined and developed during the consultation undertaken in 2014. This was supported by comprehensive documentation examining the existing characteristics of the current approach and the known issues these give rise to, and included detailed assessments of the options available for addressing them.

The consultation documents are still available to download on the issuing authority's web pages[5] and provide a detailed background to the drivers and justification for making the planned changes. It is strongly recommended that these are read in full to aid in a thorough understanding. However an overview is provided in the following section, consisting of an examination of the existing characteristics of Organisation data, the issues these give rise to, and the outline benefits of the planned changes.

[5] http://systems.digital.nhs.uk/data/ods/interfacechanges/consultation

4 Change Overview

4.1 Summary: What Are We Changing?

This standard substantially redevelops various aspects of Organisation Reference Data, governed by existing information standard SCCI0090[6]:

- move from identifiers structured with multiple-frames that indicate the type of Organisation, to a new unified structure with no inherent meaning, to ensure a sufficient number of codes for future use;
- modernisation of the delivery mechanism to adopt Extensible Markup Language (XML) instead of CSV, and general enhancement of the data model to enable:
    - greater volume of meta-data generally;
    - more complex relationships;
    - additional detail on Organisation function and remit;
    - complete date-based information;
    - context for relationships;
    - incorporation of metadata embedded in the information;
    - single, standardised file specification;
- more frequent updates (weekly at minimum with options being explored to provide access to 'real-time' updates in the future);
- county values in addresses change from mandatory to optional

4.2 Current Approach

- multiple-frame identifier structure is used for identifiers to denote organisation type (referred to as a 'Code-Frame'). E.g. NHS Trust codes are 3 characters long and begin with an 'R', CCG codes are three characters long and have a number as the first character, and letters for the second and third. There are over 80 distinct code frames currently;
- code frames also convey organisation-to-site relationships (i.e. a site identifier shares the parent organisation's identifier as a prefix);
- codes for most key organisations have historically been limited to 3 characters in length;
- flat-structure Comma Separated Value files are used to release data, with file specifications that must be separately interpreted. These flat-structure release files also mean:
    - relationships are restricted in number;
    - additional meta-data for organisations (subtype) is limited to one per organisation;
- very little information is supplied on Relationship context;
- publication is structured around quarterly 'full' releases, supplemented with monthly and weekly releases of a subset of files only;
- county element of addresses is a mandatory field; it is guaranteed to be supplied for all records

[6] http://digital.nhs.uk/isce/publication/isn/2014

4.3 Future Approach

- unified identifier structure of 5 alternating alphabetic and numeric characters (ANANA), with no inherent meaning;
- new meta-data for Primary Role and non-Primary Roles (encompassing Organisation Type and Subtype concepts respectively);
- new meta-data for Relationships providing context via a set of Relationship Names;
- increased release frequency and simplified structure;
- move from flat-file CSV format, to XML. This facilitates:
- all supporting information required to interpret data integral to file releases;
- capacity to publish multiple relationships for a single Organisation, with new meta-data that provides context for each one ('Relationship Name');
- capacity to publish multiple non-Primary Roles for a single Organisation;
- exploits opportunities for extensibility employing recognised standards-based approaches e.g. dates, OIDs, instance identifiers;
- county element of addresses is an optional; it will cease to be populated for new records in the near future

4.4 Why Are We Changing?

The changes that this submission makes to Organisation Reference Data have been designed to address a series of known issues:

4.4.1 Code Exhaustion

- code frame approach means codes are issued in discreet ranges with a finite number (particularly limited where length is restricted to 3 characters);
- approach of issuing 3 character identifiers for key organisation types has led many systems to be hard-coded accordingly. New Code Frames therefore have to be restricted to 3 characters to minimise implementation impact and this further limits total number of codes available in the range;
- combinations of codes become ‘locked off’ as a whole range is pre-assigned to a particular organisation type, but then only a small proportion is ever used;
- available number of 3 character ranges with unique formatting and a useful number of combinations is close to exhaustion; several are forecast to breach sensible limits by around 2020. A good analogy for this situation is when BT phone numbers were running out and the numbering system had to change in order to create new numbers

The new identifier structure will ensure the longevity of Organisation data by avoiding code exhaustion. It provides a total of 1.76 million codes. A total of 191,000 codes have been issued over the last 25 years (up to June 2016) so this indicates that the new structure should provide enough to last for a considerable time, even allowing for a significant increase in the volume needed.

The new approach will also prevent suppliers from hard-coding restrictions around code formats/lengths and, as a result, being unable to absorb dramatic changes to the organisation landscape. Note that any existing validation or verification prior to

input to systems that relies on identifier structure, can be replaced with the use of meta-data values.

4.4.2 Misinterpretation & Ambiguity

- There is significant risk of errors being made by consumers, and of interpretations differing from system to system:
    - Organisation type is derived from an identifier's structure rather than an explicit value, which is easy to misinterpret;
    - The information provided on relationship context is limited and differs from file-to-file (e.g. 'parent' means commissioner in one file but geographic aggregator in another);
    - Meta-data for multiple files must be manually interpreted from separate web pages or Word document specifications.

The introduction of new meta-data for Primary and non-Primary Roles, and for Relationship Names, will assist interoperability by removing all uncertainty from the interpretation of organisation data, and decreasing instances where type, relationship and subtype are misinterpreted or differ from system to system.

4.4.3 Burden of Change and Maintenance

- Effort is currently required on the part of users (for the customisation of import routines/validation) and the issuing authority (for maintenance and publication):
    - Multiple-format identifier structures mean every new Organisation type requires a new Code Frame to be defined[7];
    - Hard coding the relationship between organisation and site means a merge or split of organisations needs new site codes, even though the locations have not physically changed;
    - Multiple files are published, each one with a dedicated file description (approximately 54 files and 50 unique specifications at time of writing);
- Holding type, relationships and categories integral to the identifier or the organisation record means changing these equates to a specification change, requiring a formal change notice.

Moving from flat-file CSV format to XML will significantly decrease the day-to-day burden associated with maintaining and using Organisation data: on the issuing authority by removing the necessity for new code frames to be designed; and on consumers by removing the need to interpret increasingly complex and large numbers of Code Frames.

The use of a single standard XML schema to publish all Organisation data will also consolidate and reduce the volume of products and supporting information required, removing the current variation across products. The need to interrogate numerous disparate specifications and other supporting information will be removed as all supporting information required to interpret data will be integral to the XML file releases.

[7] This also means the issuing authority must attempt to predict the total number of instances of an Organisation type that will exist over its entire lifetime, to ensure the Code Frame can provide a sufficient volume of identifiers.

4.4.4 Inflexibility

- The current data is incapable of reflecting the increasing complexity and rate of change in the modern NHS and Social Care System, restricting innovation:
    - Use of flat CSV means the number of relationships and categories that can be provided is limited by the number of available fields;
    - Validation and import depend on the record structure remaining constant so there is no flexibility to quickly support additional attributes;
- CSV files offer limited support to emerging technologies such as Linked Data[8] and the benefits of the semantic web[9].

Adoption of XML as the primary release format provides the capacity to publish multiple Relationships, and multiple non-Primary Roles for a single Organisation. Removing the existing restrictions on relationships and categories in this way will greatly increase flexibility, making data capable of reflecting the complexity of the NHS and supporting the rate of change.

In addition, the added flexibility and extensibility of the new data model and XML release format will position the Issuing Authority to allow future support for emerging technologies such as those mentioned above.

4.4.5 Quality & Richness of Data

- current release schedules make lead times unavoidable when producing files, resulting in some changes being up to 3 months out of date at publication;
- release schedules and the current production process mean suppliers must use codes which have been allocated but not yet published;
- time-based information is limited because there is only room in the record for one set of dates (used for open/close dates on the organisation record):
    - relationships don't include start and end dates and a limited set of relationships are provided (usually the parent and its ancestors);
    - change to subtypes of an organisation over time cannot be conveyed at all;
- standardisation is prevented by the static record structure; every file has fields that have not been populated for many years or are "null" (blank)

The new approach will improve the currency of data through increasing release frequencies, and eventually reducing (potentially almost eliminating) lead times between data being changed at source and being published.

The data will also become richer through the provision of a greater volume and detail of time-based information, with the addition of an amendment timestamp and additional open/close and start/end dates across a number of elements. New meta-data for 'Relationship Name' will provide context for each relationship.

[8] https://www.w3.org/standards/semanticweb/data

[9] https://www.w3.org/standards/semanticweb/

4.4.6 Optional County Field

The legacy CSV products currently include a postal county in addresses, published as Address Line 5. In the majority of these files the value is mandatory.

This county data is sourced from address lookup software which uses the Royal Mail Postcode Address File (PAF) and is utilised by the issuing authority to validate address attributes within Organisation data.

However the postal county no longer has a valid use; it identifies postal subdivisions that ceased to be routinely use by the Royal Mail in May 1996, and they have not been required for postal purposes since.

Whilst it has remained possible until recently to continue validating the appropriate postal county value when creating addresses, this facility is due to be withdrawn. The issuing authority must upgrade to the latest industry standard address lookup data source in order to retain the ability to source accurate address information, and to incorporate the Unique Property Reference Number (UPRN) into its data (see the XML schema for details on where this will be held). Due to the withdrawal of postal county from active use, the latest address lookup products do not include it as a value that can be included in addresses.

The result is that postal county will no longer be a mandatory field in any products (including the legacy CSV files) from the implementation start date for this standard, of 24 February 2017. Shortly after this point new records will not be assigned a county value at the point they are set up.

Note that existing records that currently exhibit a county value will retain it; no steps are being taken to remove the value from historic data.

4.5 Outline Benefits

Benefit	Avoided Risk	Metric	Baseline	Expected Impact
Increased number of identifiers	Elimination of risk of identifiers being exhausted causing an inability to allocate new codes to capture new Organisations	Number of viable identifiers remaining	Remaining viable identifiers within at-risk Code Frames as of June 2016	Successful transition to new identifier structure in April 2020, giving a total of 1.76 million possible codes, will eliminate risk of codes being exhausted
Removal of necessity for new Code Frames	Reduces risk of Code Frames becoming increasingly complex in order to retain intelligence in identifiers without increasing length. Code Frames require considerable effort on the part of the issuing authority to create and maintain, and of users to correctly interpret.	Number of individual Code Frames in use, and the number of these that make use of more than one fixed character in their structure	Total number of Code Frames, and number that use multiple allowable values in a single fixed character position, as at 1 April 2016	No new frames introduced or existing ones amended following successful transition to new identifier structure in April 2020
Reducing the number of data files and supporting information products through consolidation and introduction of standard structure within XML	Reduces risk of confusion caused by an increasingly large number of separate release files with no standard structure and individual accompanying specifications for each one	Total number of release files, including supporting information such as file specifications	Number of individual products as at 1 April 2016	Reduction by one third of total number of available products following deprecation of CSV files in April 2021
Removal of ambiguity from data	Reduced risk of type, relationship and subtype being interpreted incorrectly, due to unambiguous meta-data in XML products which will provide unique identifiers and definitions for these concepts	Number of consumers taking XML products (and therefore using Role and Relationship meta-data) vs those taking CSV products	Number of downloads of XML vs. number of downloads of legacy CSV packs, from the first release following Full stage approval	Increase in number of XML downloads with corresponding reduction in CSV downloads. Split should be at least 80/20 in favour of the XML by the time the identifier structure switches in April 2020

Benefit	Avoided Risk	Metric	Baseline	Expected Impact
Improved currency of data through increased release frequencies	Reduced risk of lead times between data being changed at source and being published, causing systems to be noticeably out of date.	Number of second line support call logs received that highlight out of date information in Organisation data	Number of logs received regarding out of date information between first release following Full stage approval, and the move from quarterly/monthly to weekly in November 2017	Reduction in logs received regarding out of date information
Improved quality of data through richer information	Reduced risk of the issuing authority being unable to meet new requirements due to an inability of the model and/or release format to incorporate the data	Capture and representation of data that would have been impossible with the legacy approach; including demonstrable use of it in a live system or process	Any use of new components in the data model will be documented: Additional Attributes, multiple relationships, Other Organisations	At least one usage of each of the new components will be in evidence by the time the legacy CSV products are withdrawn in 2021
Better position to support emerging technologies such as Linked Data and the semantic web.	Reduces the risk that the issuing authority will be unable to support a new stakeholder requirement	Introduction of a new technology to support maintenance and release of Organisation data	Any use of new technology in maintenance and release of Organisation data will be documented	At least one new technology will be in evidence by the time the legacy CSV products are withdrawn in 2021

Table 1 – Outline Benefits

4.6 Alternative Approaches

Additional weight is added to the justification through an examination of the impacts of not making the change, and of possible alternative options:

4.6.1 Do Nothing

The existing interface would remain as it currently stands. All issues with the current approach would persist and many of these would worsen over time and affect stakeholders as follows:

- exhaustion of available codes; if multiple-format identifier structures are not addressed it will eventually become impossible to introduce new identifiers for new organisations within the accepted existing Code Frames;
- inability to improve on quality, currency and richness of data;
- impossible to absorb/reflect future change to the NHS;
- unable to accurately represent the true structure of health and social care;
- more difficult to support emerging technologies such as Linked Data

4.6.2 Adapt Current

The existing interface would retain use of length-restricted identifiers with a continued use of Code Frames, to avoid impacts on systems with a dependency on these existing characteristics.

In doing this however, Code Frames would have to be altered and become increasingly complex in order to retain a unique structure, as the available combinations within the three-character overall length become used up. Relatively simple Code Frames that currently include a single fixed leading character would need to be changed to allow a range of values. Indeed, this has already been done for Independent Sector Healthcare Providers; the number of codes available in their frame had to be unexpectedly increased due to expansion in independent sector's engagement with the NHS in 2011, changing from a single leading character of 'N' on a 3-character long code, to a range of 14 allowed leading characters. Detail for this change is still available in the archived information standards materials[10].

One of the key justifications for retaining Code Frames is the ability for human users to instantly recognise an Organisation's type using tacit knowledge of the structure of the code. However, as described above, the very act of persisting with the Code Frame approach will itself diminish this benefit as Code Frames become progressively more difficult to memorise.

This approach also only addresses the code exhaustion issue. The issues around Misinterpretation, Ambiguity, Inflexibility, and Quality and Richness of Data would all persist. The burden associated with use of Code Frames (on the issuing authority to maintain them, and on users to interpret them) would also arguably increase as they become more complex.

[10] http://webarchive.nationalarchives.gov.uk/+/http://www.isb.nhs.uk/documents/isb-0090/amd-36-2011/index_html

4.6.3 Other Data Sources

One more option that theoretically exists is to cease using the existing Organisation data altogether and source the information required from elsewhere. No suitable alternative sources are known, however even assuming they could be found a number of more general issues exist:

Business rules and data quality:

Data from other sources may appear to be suitable on the surface but its maintenance is highly likely to be driven by different requirements and business rules than those used currently (and upon which existing systems are reliant).

A good illustration of this is the healthcare data maintained by the Office for National Statistics[11] (ONS). This does, at first glance, appear to map across to a subset of the Health and Social Care Organisation Reference Data exactly. However, ONS in fact identify the geographic boundaries for NHS entities, rather than the organisational entity. This is a subtle but key difference that would cause significant issues if the ONS codes were to be implemented as organisation reference data in clinical systems. This is because ONS issue a new code whenever the *boundary* for an organisation is changed, regardless of whether the organisation entity itself is affected.

Scope of Data:

Numerous additions and extensions to Organisation data have been made over the years to capture new organisation types, in order to cater for new policy developments. This has led to a very wide range of organisation types being encompassed, extending outside the scope of what can be considered standard NHS entities (e.g. schools, prisons, local authorities etc.). It may be possible to uplift another dataset to encompass the key NHS organisations however it is extremely unlikely that any other source will cover the breadth of organisation types that the existing data does, or have the capability and flexibility to meet ongoing requirements for the capture of data about new types of entity.

Fitness for Purpose:

Countless variations and tailored elements of Organisation data have been produced over the years to cater for specific system and application requirements – such as the Legal and Operational dates that the XML will provide. Any other source of organisation information, no matter how extensive in terms of scope, is highly unlikely to allow for these.

Burden of Change:

If an alternative data set could be found that follows similar business rules, captures the full scope of organisation types and can be made to fit all the requirements of current users, a significant additional impact will result from the need to replace all current codes used in systems with those from the alternative source. Arguably a greater impact will result from this than the changes this paper outlines.

[11] www.ons.gov.uk

5 Costs

For the purposes of this change, costs can be split into two distinct categories:

- Some unique costs will be incurred internally, by the issuing authority responsible for collating, maintaining and publishing Organisation data, in order to change infrastructure and processes and enable production of the new interface;
- The majority of costs will be incurred externally, by stakeholders of Organisation reference data in making changes to the systems and business processes that use it.

5.1 Internal

At the time of writing, the issuing authority responsible for Health and Social Care Organisation Reference Data is the Organisation Data Service (ODS), part of Systems and Services Delivery within the Operations and Assurance Services directorate of NHS Digital.

A considerable volume of work has already been completed by ODS to define the specification for the new approach, which has included analysis of known issues, some formal consideration of various options and a period of open consultation with users, as well as the development of this standard and the technical specifications for the new data model and XML release products. All of the resources that have been needed to date have been covered within the ODS business as usual budget.

Further work will be required to test and implement all the internal processes, systems and tooling required to enable the new XML outputs to be produced, generate all the necessary communications, provide support and query resolution to stakeholders in their transition to the new model, and to continue maintenance of the service into the future. It is not expected that this transition stage will demand any significant *new* resource from a service provision perspective however ODS is entirely funded via Grant in Aid, so if any additional funds are required they will be requested and (it is anticipated) secured via the normal annual budgeting process.

5.2 External

The breadth of use of Organisation data is huge – virtually *all* systems, applications and organisations throughout the health and social care system make some use of it, and many outside of health and social care too.

Many systems that use the data will have aspects of it hard-coded into their architecture; absorbing the changes will therefore be difficult and is likely to be expensive. Other systems will make limited use of a small amount of data, meaning they are able to absorb the changes with relatively little effort or expenditure. Or there may be no hard-coded system reliance at all – an organisation might periodically use some organisation codes in a manual or ad-hoc process that is set up from scratch, so whilst they would have to take note of the changes, they would not contribute to additional effort or cost.

So, whilst it is possible to state that *all* existing users of the data regardless of their business area *will be impacted in some way,* this will be to differing degrees and

result in wildly varying costs, depending on the nature of implementation – and because the data is publically available it is extremely difficult to exhaustively identify all uses of it in detail.

The issuing authority alone does not have the resources required to quantify whole-system costs at any level of accuracy; this will require a significant amount of specialised work focusing on this area alone.

In order to take this area of assessment further, and in response to a recommendation from the SCCI board following approval of the Requirement stage of this standard, steps are being taken to establish a dedicated programme of work focusing on the implementation of the changes to Organisation Reference Data. As this work progresses, more detail will emerge on the volume and nature of impacts across health and social care, and what they will cost.

Therefore, a fully comprehensive cost estimate for the changes is not available at the time of writing to insert into this document. However, one of the new programme's key deliverables will be to estimate this figure and details of the programme including its governance arrangements, funding and planned deliverables will be presented to the SCCI board at Full stage, so its existence (and fitness for purpose) will inform the approval of this information standard. In addition, a key requirement of the programme will be to continuously report back to SCCI following approval of the standard, so that they may consider the emerging information on costs and impact, as implementation progresses.

6 Funding

Whilst finding for internal costs incurred by the issuing authority is covered (via the business as usual budget), funds are **not** available centrally to cover costs incurred externally, to aid any end-users of Organisation data in making changes. This includes users within NHS Digital.

It is however vital that the scope and quality of the changes are not compromised. The problems with the current approach and an increasingly fluid health and social care landscape demand that the new interface is implemented as proposed, in order to provide robust data that continues to meet the needs of users for years to come.

With cost and quality constrained in this way, it is clear that the timescale for implementation is the only factor that can be adjusted to reduce impact for end-users. The implementation plan is designed accordingly by perpetuating some of the existing characteristics of the legacy products for as long as possible, thereby extending support for any dependencies on them. It is hoped that this delay will be long enough to allow end-users to make the changes required using planned annual development budgets or perhaps work them into re-procurement activities, without any need for further unplanned funding.

More detail is available within the Implementation Strategy document; however the two broad approaches that are being used to achieve this are outlined below.

6.1 Delaying Identifier Structure Transition

Although the current approach of building meaning into structured identifiers clearly risks exhausting available codes this will be continued for some time, with the transition to unstructured codes being delayed until April 2020.

This timescale is based on analysis of the remaining available code frames and the rate of current usage, which has determined that at worst case the maximum safe period of time that the legacy arrangements can be persisted without risking exhaustion of key Code Frames is between four and five years. This analysis is included in full, in the Appendix.

6.2 Dual running CSV files with XML

Although the introduction of XML format files is one of the first planned milestones (November 2015, in beta), this will not initially be at the expense of the existing CSV format files, which will continue to be supported through until the end of 2021. The two will be provided in parallel for this period.

7 Summary

The changes *must* happen. The alternative is for code allocation to become prohibitively complex to maintain in the near-to-medium future, and eventually impossible, due to code exhaustion. Exhaustion is forecast to begin to occur for some key types of Organisation by 2020. The impact of not being able to issue unique codes at all would undoubtedly outweigh the impacts of the service adopting the new approach (and this does not factor in the long-term benefit of modernisation that the changes will bring).

Some of the detailed impacts of the changes outlined by some stakeholders in the BAAS assessment serve, somewhat ironically, to strengthen the overall argument for the proposals. Although they illustrate the huge volume of change that is going to be required in some settings, this simultaneously highlights the degree of dependency on the data and the catastrophic impacts that would result from an inability to issue any new identifiers at all.

7.1 Possible Further Mitigation

The issuing authority recognises that, in light of the level of cost indicated in some of the feedback secured by BAAS, further compromise on the transition approach may be unavoidable. Whilst undesirable, it would be possible to delay the key change milestones that have already been set in the future – outlined above in sections 6.1 and 6.2 and detailed in full in the Implementation Strategy document.

It is important to caveat this proposal with recognition that doing so will bring some subsequent impacts; retaining structured identifiers for longer will increase the risk of an exhaustion point being reached in one or more of the existing Code Frames. Should this occur the issues outlined in section 4.4 will begin to have an impact, however more time can be bought through interim changes being made to at-risk Code Frames to increase the range of identifiers they encompass whilst retaining the overall length restriction. This can be achieved through the addition of extra allowable values within the fixed character positions defined within a Code Frame. There is precedent for this – see section 4.6 for an outline of the changes made to the Code Frame for Independent Sector Health Providers – however:

- The approach is not sustainable – eventually identifiers will reach total exhaustion regardless of the mitigating actions taken;
- Addressing individual coding frames isn't logical as across the board ranges of codes are exhausted at different rates;
- One of the key justifications for retaining the Code Frame approach (that they are easily recognisable for human users) will be further and further diminished if this approach is adopted across numerous code frames (also detailed in 4.6).

In light of this, the approach cannot continue indefinitely. Should the option be recommended by SCCI and adopted, a concrete transition date *must* be specified in the future. Simply continuing until exhaustion reaches a critical stage is too risky; it would create uncertainty, would not account for a sudden requirement for a large volume of codes (such as might be triggered by further reconfiguration of the NHS),

and would not provide sufficient notice to be given to users regarding the eventual implementation of the changes.

Appendix

Remaining Codes Analysis

An in-depth analysis of organisation coding frames for active organisations has been carried out for all organisation types which have three characters. The term “Active” should be interpreted as an applicable organisation coding frame which has been used to issue an organisation code in the last ten years. All those three character coding frames which do not appear in Table 2 below are either:

- **Closed** - in which case all organisations of that type are closed and no further codes will be issued from that range
- **Still In use** – but the rate of churn over the last ten years has reduced to zero

Note that the figures in Table 2 below include data collated up to June 2016. Organisation coding frames that are at particular threat are in red italics.

Organisation Type	Leading Character	Format	Max @ 80% Threshold	Used	Remaining @ 80% Threshold	Average Churn	Forecast Years @ 80% Exhaustion
SHA/Area Team/Region Geography	***Q***	*Q, 0-9, 0-9*	*80*	*79*	*1*	*5.64*	*0.14*
Cancer Network	***N***	*N, 0-9, 0-9*	*80*	*57*	*23*	*7.92*	*2.32*
NHS Trust	***R***	*R, A-9, A-9*	*1037*	*682*	*355*	*29.02*	*9.78*
Primary Care Trust [PCT]	**5**	5, A-Z, A-Z	1037	434	603	39.45	12.22
National Application Service Provider	**YE**	Y,E, A-9	29	12	17	1.04	12.88
Clinical Commissioning Group		0-9, 0-9, A-Z	2080	254	1826	84.67	17.25
Non-Statutory NHS Organisation	**YA**	Y, A, 0-9	8	2	6	0.27	17.52

Organisation Type	Leading Character	Format	Max @ 80% Threshold	Used	Remaining @ 80% Threshold	Average Churn	Forecast Years @ 80% Exhaustion
Health Authority	**Q**	Q, A-Z, A-9	749	111	638	22.20	22.98
IT Cluster	**Y**	Y, 0-9, 0-9	80	29	51	1.49	27.43
Commissioning Support Unit	**0**	0, A-Z, A-Z	541	35	506	11.67	34.68
Executive Agency	**X**	X, 0-9, 0-9	80	6	74	0.71	83.87
Data Management and Integration Centre	**0**	0, A-Z, A-Z	541	11	530	3.44	123.30
Care Trust	**T**	T, A-Z, A-Z	541	12	529	1.50	282.03
Government Department	**X**	X, A-Z, A-Z	541	10	531	0.40	1061.60

Table 2 - Analysis of active three character organisation types[12]

The key piece of data in the table is in the “Forecast 80% Exhaustion” column. The entries in the table have been sorted in ascending order, so those with a shorter forecast life span appear first.

The consultation materials stated that unstructured identifiers would be introduced at the point at which coding frames for key types of organisations are forecasted to be 80% exhausted. The organisation type used to represent SHA/Area Team/Region Geography organisations is central to the operation of Health and Social Care and based on past history there are sufficient codes available to support code allocation for less than 12 months based on the 80% threshold (therefore 20 available codes remain). However the churn rate of 5.64 codes per year is an average applied since SHAs were introduced in 2002. In more recent years the churn has increased following the introduction Area Teams as part of the Health and Social Care Act and further reconfiguration activity to form Region Geographies. More detail is provided in Figure 1 below.

[12] The figures presented in Table 1 are all based on 26 characters where A-Z is shown in the Format column and 36 characters for A-9. The recalculation has been amended to take into account the decision to reintroduce the previously illegal range of B,I,O, S,U and Z in the Unified Identifier Structure.

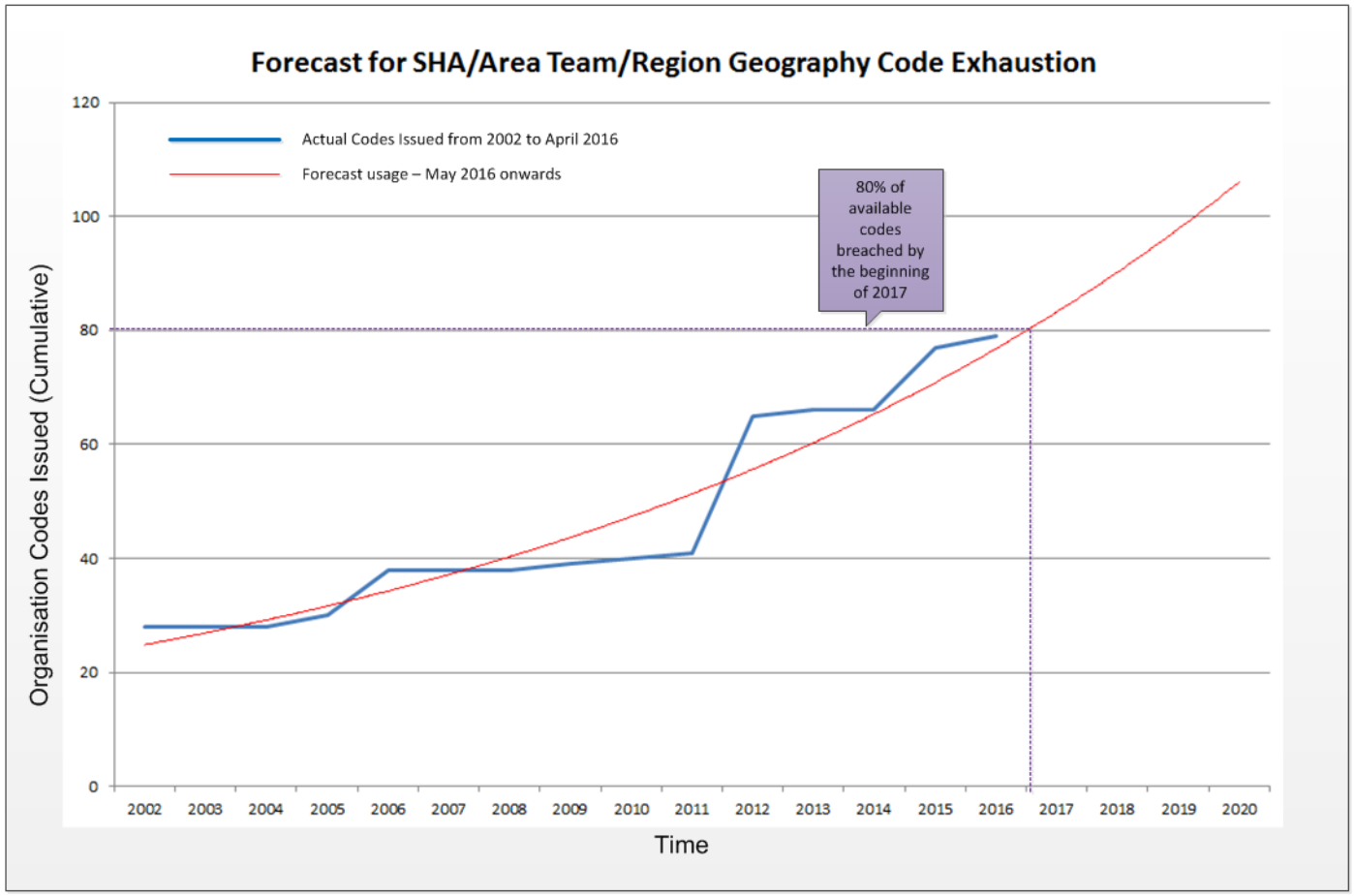

Figure 1 - Forecast of SHA/Area Team/Region Geography Code Allocation

Region Geography codes have been forecast to be completely exhausted by mid-2019 in advance of the transition to unified structure identifiers. Fortunately, there is a workaround in this particular situation which permits allocation of Region Geography codes beyond 2019; since 2002 SHA/Area Team/Region Geography organisations have been issued using a three character coding frame of Q followed by two digits (i.e. **Q**, 0-9, 0-9) as shown in Table 2, e.g. Q35. "Q-codes" of this format have been used

throughout Health and Social Care to represent High Level Health Geographies[13] for approximately 16 years. Prior to the introduction of SHAs however, Q-codes were used to represent Health Authorities but the coding frame differed (**Q**, A-Z, A-9) as shown in Table 2, e.g. QA3, QAH.

An updated approach for Q-code allocation will be put in place which involves issuing codes to represent High Level Health Geographies from this expanded range.

Whilst this approach buys more time and could potentially be applied in other at-risk code frames, this demonstrates that the types of issues discussed in section 4.6.2 (adjustment of coding frames) are real and would continue to impact on code allocation if the option to adapt the current coding approach was adopted.

Note that the last code that can be issued in the current range (**Q**, 0-9, 0-9) is Q98; this is because Q99 has already been allocated as a default code for local systems. Once Q98 has been allocated codes will be issued from available pockets of codes including the range previously used to allocate Health Authorities – please see Table 3 below:

Organisation Type	Leading Character	Format	Max @ 80% Threshold	Used	Remaining @ 80% Threshold	Average Churn	Forecast Years @ 80% Exhaustion
SHA/Area Team/Region Geography	Q	**Q**, 0-9, 0-9	80	79	1	5.64	0.14
Health Authority	Q	**Q**, A-Z, A-9	749	111	638	22.20	22.98
New Q-Code Range	Q	**Q**, A-9, A-9[14]	1037[15]	190[16]	846	5.64[17]	150.14

Table 3 - Planned changes to Q-code format

[13] The term “High Level Health Geography” was introduced following the Health and Social Care Act to help insulate standards from change

[14] Provides 1296 codes in total (36 x 36)

[15] 80% of 1296

[16] Sum of Q-codes used in current ranges

[17] Based on the average churn since the introduction of SHAs in 2002

Fortunately the proposed Q-code range does comply with the coding frame published in the Data Dictionary (see Figure 2 below) and for this reason an Information Standards Notice will not be required. Despite this, there is still some risk attached to extending the range as systems (particularly those which have been introduced after 2002) may be expecting Q-codes to continue to only have digits in both position 2 and position 3. During the intervening period up until mid-2019 details of the approach will be communicated to consumers to mitigate the risk.

ORGANISATION CODES TABLES

Table 1: CODING FORMATS FOR ORGANISATIONS IN ENGLAND AND WALES

Organisation Type	Frame Type				
	See Coding Frames Table	1	2	3	
High Level Health Geography, e.g. NHS England Region (Geography)	D	Q	A-9	A-9	

Figure 2 - High Level Geography coding frame in the NHS Data Dictionary[18]

[18] Source: http://www.datadictionary.nhs.uk/data_dictionary/attributes/o/org/organisation_code_de.asp?query=coding%20frame&rank=1&shownav=1