Ref: care.data/Programme Board/280114_04

Title:
Proposed approach for selecting which additional patient-level data sets should be added to phase 1 of the care.data programme.

Author: Peter Flynn, Programme Director – care.data, NHS England

Programme Board Sponsor(s): Chief Data Officer, NHS England

Purpose:
To inform and gain support from the care.data programme board for the proposed process for selecting which additional patient-level data sets will be incorporated, and made available as linked data, in the first phase of the programme (to the end of FY2015/16).

Background:
There are currently at least a further 23 identified data sets that could be included. These data sets differ in size, complexity and the ease with which they could be added. Once the Strategic Outline Case (SOC) is approved for the programme, we need to be able to proceed to develop detailed benefits statements for the Outline Business Cases(s).

Key Points:
NHS England, as lead commissioner, will be making a recommendation to the programme board as regards which data sets should prioritised by the programme during phase 1. It is expected that this recommendation will be no later than the May 2014 board meeting.

We plan to base our recommendation on policy commitments, an analysis of (a) the anticipated **benefits** of linking each data set, and (b) the **availability** of each data set.

We will conduct this analysis in consultation with nominated individuals from each ISCG partner organisation.

Desired outcome(s):
That the board:
1) approve this approach to generating a recommendation; and
2) nominate individuals to represent their organisation as part of the analysis.

Background

1. The approach to the delivery of the care.data programme is to upgrade the hospital episodes statistics (HES) service into a new service; so that it incorporates:
    a. missing details of the care provided in hospital; and
    b. missing details of the care provided outside hospital.

    The working title, within NHS England, for this new service is the care episodes service (CES).

2. Accordingly, the current priorities of the care.data programme are (a) implementing the findings of the hospital data consultation and (b) the incorporation of data from GP practices. These priorities will continue through 2014/15.

3. The draft Strategic Outline Case (SOC) for care.data includes the capacity to incorporate an additional 17 datasets of average complexity and size over phase 1 of the programme (phase scheduled to run to the end of March 2016).

4. We have identified 24 data sets that could be included in phase 1 (see Box 1).

Box 1: Candidate data sets for inclusion in phase 1 of the programme

a. Ambulance	i. GP Prescriptions (Dispensing)	q. Pathology (GP requested: results of investigations)
b. Births	j. Hospital activity (PbR)	r. PROMs/PREMs
c. Child & Adolescent Mental Health Services (CAMHS)	k. Maternity	s. Psychological therapies (inc. IAPT)
d. Children's	l. Mental health[1]	t. Referrals
e. Community information	m. National clinical audits	u. Registries
f. Deaths (ONS mortality data	n. NHS health checks	v. Screening
g. Diagnostic Imaging Data (extension)	o. Other primary care	w. Social Care (Adult)
h. Genomic data (from volunteers)	p. Out-of-hours	x. Systemic Anti-Cancer Therapies (SACT)

5. These data sets (in Box 1) differ in size and in the degree to which they are already effectively 'available' (i.e. the extent to which they are already defined, standardised, collected electronically, and flowing to the HSCIC).

6. The collection and distribution of Genomic data, albeit from volunteers, may be sufficiently challenging that it warrants a completely separate business case from the current care.data case (SOC).

---

[1] Mental Health Minimum data set (MHMDS) is already available linked to HES data.

Proposal

7. The programme board will be asked to decide which of the data sets (in Box 1) should be prioritised for phase 1 of the care.data programme.

8. We propose making our recommendation to the board based on an analysis of:
   - Policy commitments
   - The anticipated high level **benefits** of incorporating each data set
   - The **availability** of each data set, as defined above

   This will incorporate work that is already underway, e.g. Child and Maternity health data.

9. As highlighted in the Board Report, the approach to benefits planning and realisation for the programme as a whole is underway. This work will deliver a comprehensive statement of the anticipated benefits of the programme for patients and citizens.

   However, in the short term, we need to conduct a rapid assessment of the high level benefits of each candidate data set, in order to make decisions about which data sets should be prioritised in phase 1. This will enable early development work to take place to inform the Outline Business Case (OBC).

10. We therefore propose conducting an assessment of high level **benefits** of holding and linking each candidate data set in consultation with nominated individuals from each ISCG partner organisations. The assessment will focus on the anticipated benefits to:
    - quality
      - patient safety
      - patient experience
      - effectiveness
   - efficiency
    - equity

    This analysis will be used to generate an aggregate relative benefit score for each candidate data set.

11. Using the same approach we will assess the **availability** of each candidate data set in terms of it being:
    - already defined (i.e. an agreed dataset approved by ISB/SCCI)
    - standardised (i.e. it uses a terminology standard, such as SNOMED/Read/ICD10)
    - collected electronically
    - flowing to the HSCIC

    Again, this analysis will be used to generate an aggregate relative availability score for each candidate data set.

12. Based on its aggregate benefit score and aggregate availability score, each data set would then be positioned on a 'Boston matrix' of benefits versus availability (see Figure 1).

Figure 1: Illustrative Boston matrix of benefits versus availability of each candidate data set
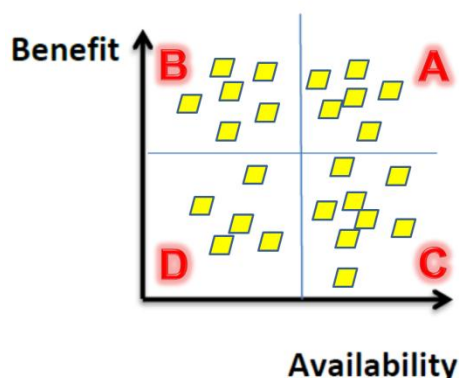
13. The data sets in area A (high benefits, high availability) would be considered as "accelerator projects" for rapid incorporation into CES. In this context an "accelerator project" would include a scoped project focused on a small geography or a small number of data suppliers to inform a national collection, e.g. working with a small number of trusts who already collect prescribing data electronically.

    The data sets in area B (high benefits, low availability) would be considered as "slow burner" projects requiring more technical consideration and support to bring into CES. There may be candidate "accelerator projects" here too.

    Data sets in area C (low benefits, high availability) would be could be used either as "low hanging fruit" (i.e., to bring in quickly and early), or as opportunities to maximise the use of available resources whilst focussing on data sets in quadrants A and B.

    Data sets in area D (low benefits, low availability) should be de-prioritised in the short term.

Recommendations

1) The board is requested to approve this approach to identifying benefits and availability of data sets, in order that the delivery plan for phase 1 of the care.data programme, in line with the Outline Business Case covering this phase, can be agreed.

2) Board members are asked to nominate individuals to represent their organisation's view of the benefits of having linking each data set, at patient level, into CES.

Peter Flynn
Programme Director – care.data
NHS England

23 January 2014