

# Department for **Education**

**Data and Statistics Division**

**2012 Year 1 Phonics Screening Check Data Collection**

**Technical Specification**

**Version 1.0**

---

## Contents

# Department for Education

	1
1. Introduction	3
1.1 Audience and Purpose	3
1.2 Scope	3
1.3 Business Rationale	4
1.4 Assumptions	4
1.5 Coverage and Timings	5
1.6 Data Standards	5
2. Structure and Guidance	6
2.1 Outline data content	6
2.2 Process	6
2.2.1 Entry of marks and outcomes into school MIS	7
2.2.2 MIS data entry checks	7
2.2.3 Creation of CTF phonics file for transmission to LA	8
2.2.4 Loading Phonics results into the LA's central management system	8
2.2.5 Creation of individual pupil results XML export file by the LA or Accredited Agency for transmission to the DfE	8
2.2.6 Loading Phonics results into the LA's central management system	9
2.3 Additional LA software facilities	9
2.4 Reports for schools and parents	10
2.5 Example file structure for XML returns	11
3. Data Validation Rules	13
3.1 Header Validation	14
3.2 Establishment Level Validation	14
3.3 Pupil Level Validation	14
4. Resubmission Guidance	17
5. Valid values	17
5.1 LA codes	17
5.2 Establishment Types	17

## Version History:

0.1	Initial draft	Terance Marion 09 <sup>th</sup> June 11
0.2	Second draft incorporating comments from Adam Hatton and Anthony Agbueze	Terance Marion 11 <sup>th</sup> July 11
0.3	Third draft incorporating comments from Rob Bauling	Terance Marion 13 <sup>th</sup> July 11
0.4	Fourth draft incorporating comments from Claire-Louise Patrick and Phil Elks	Terance Marion 19 <sup>th</sup> July 11
0.5	Changes relating to file processing by COLLECT and riders on check outcomes	Rob Bauling 9 <sup>th</sup> Aug 11
0.6	Changes requested by Adam Hatton and Phil Elks	Terance Marion 23 <sup>rd</sup> Aug 11
0.7	Changes requested by Adam Hatton Alex Miller, Phil Elks, Claire-Louise Patrick, Chris Wood and Ayo Babatunde,	Sham Malik 2 <sup>nd</sup> Feb 12
1.0	Baselined	Sham Malik 8 <sup>th</sup> Feb 12

## 1. Introduction

### 1.1 Audience and Purpose

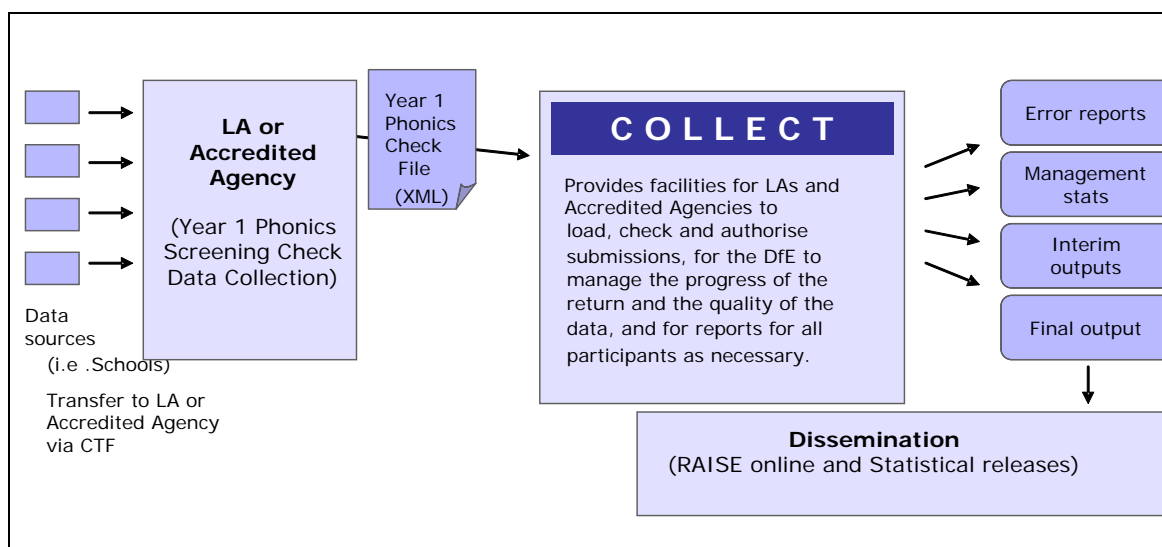
This specification describes the 2012 Year 1 Phonics Screening Check data collection. It will enable Schools, Academies (including free schools), Local Authorities (LAs) or accredited agencies, and software suppliers working on their behalf, to prepare the necessary data and processes for compliance so that all data relating to the Phonics Screening Checks carried out in 2012 can be returned to DfE.

The specification covers the expected return to the DfE and the validation rules which will be applied. It should be read in conjunction with the current CBDS definitions and 2011/12 Assessment Components (A\_COMP) Module which can be found on the Department's website.

### 1.2 Scope

This data collection process covers all Maintained Schools and Academies (including free schools) in England, submissions are processed through an LA or Accredited Agency who in turn will provide data on all Year 1 Phonics Screening Checks carried out in 2012. The data consists of pupil-level information.

The following diagram outlines the overall scope of the collection:



*Fig 1: High level data collection process*

This requirements specification consists of one document containing a narrative, sample XML and validation. LAs and suppliers should also refer to the latest version of CBDS (Common Basic Data Set) posted on the Department's website below:

<http://www.education.gov.uk/schools/adminandfinance/schooladmin/ims/datamanagement/a0058744/common-basic-data-set-cbds-database>

### 1.3 Business Rationale

The following factors are the business drivers behind this approach:

- It is consistent with the overall vision behind the Data Sharing Protocol of collecting data once and using it many times;
- It is based on the principle that local authorities and schools are expected to hold/manage data for their own purposes, or have data held on their behalf so that they can access and extract it.

### 1.4 Assumptions

Assumptions made in creating this specification:

- A1** Where appropriate the data specification has used CBDS/ISB data standards
- A2** The Return will use COLLECT as the data collection mechanism
- A3** DfE will only accept submissions in XML format from LAs and Accredited Agencies [

## **1.5 Coverage and Timings**

Maintained schools (including Academes and Free schools) will capture the data specified for the 2012 Year 1 Phonics Screening Check. The data will then be collected via DfE's COLLECT system between 18th June and 31st July 2012. LAs and accredited agencies may be able to correct errors until 30th September 2012. Data received after 31st July 2012 may not be included in the provisional statistics or release of data in RAISE online.

## **1.6 Data Standards**

It is the long-term strategy of the Department to adopt the ISB standards to support the transfer of data across other public sector bodies and government departments. This specification has been developed with those standards in mind; where deviations occur explanation notes have been provided. Details of the ISB standards can be found here - <http://www.education.gov.uk/escs-isb>

## 2. Structure and Guidance

### 2.1 Outline data content

Each return from a LA and accredited agency will consist of one or more files each of which has a header and a school module containing a pupil module with a number of pupil records. Each pupil record will include identifiers plus the mark and outcome achieved in the Year 1 Phonics Screening Check. Multiple files should be returned as a .zip file.

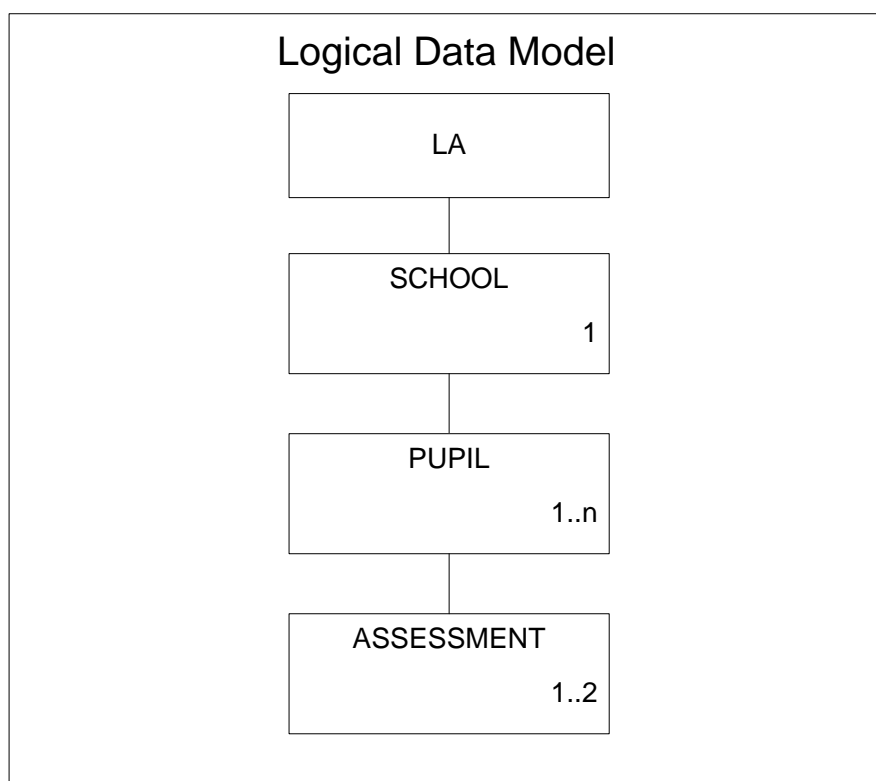


Figure 2 - Outline structure of data.

(Note: Replace LA for agent or other accredited agent and School for Academy or Free School)

### 2.2 Process

The process will involve an annual screening check where teachers will assess students on their phonics decoding ability. Students will be asked to read out 40 items with a teacher. These items will be a mix of real words and pseudo-words. The teacher will then score each item (1 mark for a correct answer) and total the scores to make up the mark for the screening check. Based on this mark the student will be assigned an outcome. Teachers will be told the threshold score for the check (marks out of 40 required to meet the standard) when they open the materials for the assessment.

By the end of the assessment the teacher should have information to populate two fields, mark and outcome, for each student. This can also be automated and MI systems could calculate the Outcome based on Mark and Threshold supplied.

**Monitoring:**

Local authorities will monitor 10% of the schools in their area. Monitoring visits will take place before the check (to check materials are being held securely), and during the time of the check to ensure the administration process is being followed and observe some screening checks taking place. If necessary, local authorities can also carry out monitoring visits immediately after the check week. LAs have been informed of the general arrangements, and the funding being provided for this work. A guidance document will be sent out for monitors to LAs in the Spring term of 2012.

**2.2.1 Entry of marks and outcomes into school MIS**

In addition to the basic pupil identifiers (as collected in other collections), there will be 4 data items for each pupil.

The four data items 'Phonics Outcome', 'Phonics Mark', 'Phonics Recheck Outcome' and 'Phonics Recheck Mark' are the required components of the new 'Year 1 Phonics Screening Check' data collection (but note the recheck items will not be collected until 2013).

The 'Phonics Outcome' and the 'Phonics Recheck Outcome' fields are to be populated by the outcome of a Phonics Screening Check. These will be populated by selection from a codeset with the following values:

- Wa – met the expected phonic decoding standard for a pupil at the end of year 1
- Wt – has not met the expected phonic decoding standard for a pupil at the end of year 1
- A – Absent
- Q – Maladministration
- D – Disapplied
- L – Left

The 'Phonics Mark' and the 'Phonics Recheck Mark' will record the actual mark from the screening check (currently expected to be a mark out of 40) as a two digit integer. There will be specified mark thresholds which then determine the outcome. Pupils who do not meet the expected standard in the phonics check at the end of Year 1 will retake the check in June 2013. Results from the recheck will also be collected by the Department, using the Collect system.

**2.2.2 MIS data entry checks**

The school MIS software should check that a valid entry has been made in the fields for the Year 1 Phonics Screening Check where appropriate. Any invalid or missing entries should be reported to the school when the data is entered and/or when the CTF is to be created.

### **2.2.3 Creation of CTF phonics file for transmission to LA**

Maintained Schools must submit the results of the Year 1 Phonics Screening Check to their LA. Academies (including free schools) must submit the results of the Year 1 Phonics Screening Check to their Geographical LA or an Accredited Agency. This will normally be via a CTF and when using this route, the name of the CTF created by schools will be:

xxxnnnn\_CTF\_xxxLLLL\_yyy.XML where xxx represents the LA number, nnnn is the school number and yyy is an incremented version number starting at 001.

### **2.2.4 Loading Phonics results into the LA's central management system**

- i) Where any pupil identifying or contextual data are already present in the LA central database there should be a prompt to ask if the existing data should be overwritten by the CTF or CSV import.
- ii) Where a CTF contains any pupil's previous assessment data (e.g. earlier year's EYFSP result) there should be a prompt to ask if the existing assessment data should be overwritten.
- iii) The LA should also have facilities to:
  - to remove and to add pupils;
  - to amend a pupil's identifying data or result levels that have been loaded, and for the new data to be re-validated in accordance with the rules in sections 4.1 to 4.3.
  - to re-import a corrected CTF with an option for the operator to either replace all existing data or add as new data, and be validated in accordance with the rules in sections 4.1 to 4.3.

### **2.2.5 Creation of individual pupil results XML export file by the LA or Accredited Agency for transmission to the DfE**

- i) On completion of the actions at 2.2.4, the software should allow the LA or the Accredited Agency to create separate XML files of individual school and pupil data, as specified in the example XML file at section 3.2, even if still containing errors, for loading into the DfE data collection system (COLLECT). Where no assessment Result level is reported for a subject, all Assessment components for that subject should be excluded from the XML file. The XML file should be validated in accordance with the rules in sections 4.1 to 4.3. Any data failing the validation checks should be reported using the error codes and messages specified. In addition to LA and school number, the file will contain the following data items for each pupil:



UPN  
pupil surname  
pupil forename  
DoB  
Gender  
Y1 Phonics Check Result - Mark as recorded  
Y1 Phonics Check Result – Outcome as recorded

**Please note** – An approved SoftwareCode will not be issued. COLLECT will use the SoftwareCode tag value specified by suppliers for this purpose.

- ii) The convention for naming the XML file is:  
xxxLLLL\_Y1P\_xxxDfE\_yyy.XML where xxx represents the LA number, and yyy is an incremented version number starting at 001.
- The LA should be provided with options to include in the export:
    - a) **all** schools' files even if these contain errors or have previously been exported (- the software should inform the operator if a school is in error or has previously been exported);
    - b) **selected** schools' files (i.e. a single school or multiple schools), even if these contain errors or have previously been exported (the software should inform the operator if a school is in error or has previously been exported).

***N.B.*** *Whichever option above is chosen, all individual pupil results within the school are to be exported. There must be one file per school*

## 2.2.6 Loading Phonics results into the LA's central management system

In addition, the software should allow the LA to load into its central database a CSV file created by schools from a Phonics Excel Independent spreadsheet (to be supplied to LAs and Accredited Agencies by the DfE for distribution if necessary) with the filename Phonics2012 DataInputToXML v1-0.xls. A copy of the spreadsheet can be provided when available, if required.

## 2.3 Additional LA software facilities

The software should provide the following additional facilities:-

- i) to allow the LA/Accredited Agency to produce a list of schools to check that all expected schools returns have been processed;

- ii) to allow the LA/Accredited Agency to produce a report for each school or for all schools to show total number of pupils included in each school's CTF submission. It could also include number of boys and girls as an additional check.
- iii) to allow the LA/Accredited Agency to transmit results in batches without the need to resubmit results already transmitted.

## 2.4 Reports for schools and parents

Schools will be expected to report to parents or persons with parental responsibility Year 1 Phonics Screening Check outcomes for their children. Further information is given in Section 5 of the Year 1 Phonics Screening Check Assessment and Reporting Arrangements document which is available on our website:

<http://www.education.gov.uk/assessment>

### Data Return Formats

A submission file from an LA/ Accredited Agency should be provided in a XML file structure as described in 3.2. Where an optional item is not included then the associated tag should also be omitted. Where a mandatory item is omitted then an error will be generated. Special Notes for XML returns  
The standard XML entity references should be used for the following special characters:

Character	Entity reference
Ampersand (&)	&amp;
Left Angle Bracket (<)	&lt;
Right Angle Bracket (>)	&gt;
Single Quote / Apostrophe (')	&apos;
Double Quotes ("")	&quot;

For special characters such as é use a character reference such as &#233; This produces a generic XML file which can be viewed as such in a suitable browser. Details of all such characters are found within the XML standard documentation; see for example <http://www.w3.org/TR/2004/REC-xml-20040204/>, section 2.2.

## 2.5 Example file structure for XML returns

CBDS Ref	XMLFormatLayout	Notes
	<?xml version="1.0" encoding="UTF-8"?>	
	<Y1PhonicsFile>	
	<Header>	
800001	<Collection>Y1 Phonics Transfer File</Collection>	Fixed value of "Y1 Phonics Transfer File"
800010	<DateTime>2012-06-23T15:30:47</DateTime>	
100266	<Year>2012</Year>	Fixed value of '2012'
200001	<LEA>302</LEA>	Valid values listed in section 6.
800006	<SoftwareCode>KEYPAS</SoftwareCode>	Suppliers decide on code used by their system
	</Header>	
	<School>	
200002	<Estab>3501</Estab>	Valid values listed in section 6.
	<Pupils>	
	<Pupil>	A repeatable group of 1 or more
100001	<UPN>A123456789012</UPN>	
100003	<Surname>Smith</Surname>	
100004	<Forename>John</Forename>	
100007	<DOB>2006-02-02</DOB>	
100008	<Gender>M</Gender>	
	<Assessments>	
	<Assessment>	
100466	<Subject>PHO</Subject>	<Assessment> : A repeatable group of 1 or more depending on the combination of <Subject>, <Method>, <Component> and <ResultQualifier> required. See A_Comp 2012.
100273	<Method>TT</Method>	
100274	<Component>CHK</Component>	
100275	<ResultQualifier>NY</ResultQualifier>	
100276	<Result>Wa</Result>	
	</Assessment>	

	<Assessment>	<Assessment> : A repeatable group of 1 or more depending on the combination of <Subject>, <Method>, <Component> and <ResultQualifier> required. See A_Comp 2012.
100466	<Subject>PHO</Subject>	
100273	<Method>TT</Method>	
100274	<Component>CHK</Component>	
100275	<ResultQualifier>NM</ResultQualifier>	
100276	<Result>39</Result>	
	</Assessment>	
	</Assessments>	
	</Pupil>	
	</Pupils>	
	</School>	
	</Y1PhonicsFile>	

### 3. Data Validation Rules

This section specifies validation rules that are applied to the data by DfE, using COLLECT. They should also be applied, as far as possible, by LA/Agents prior to loading the data onto the COLLECT system and any issues resolved. Those data items included in the specification but not appearing in the validation rules - that is, those not shown in the validation rules as "must be present" - should be supplied where available (e.g. child's UPN). The rules also show which validations are Errors and which ones are Queries. An Error is a failed validation check that must be corrected. A Query is one which must be investigated, and potentially corrected, as it identifies data that would usually be invalid.

Syntax:

- Validation checks are expressed as conditions that must be true (errors) or that should be true (queries). The Error Message shows what is displayed if the condition fails
- data items are indicated using the XML tag with the CBDS number in parentheses, for example <UPN> (100001)
- data groups are indicated using the XML tag only, for example <Assessments>
- single dates or date ranges are sometimes defined in relation to <ReferenceDate>, which is from the XML Header.

### 3.1 Header Validation

Seq Number	Error/Query	Validation Check	Error Message
001	Error	<Collection> (800001) must be present with a value of "Y1 Phonics Transfer File"	Document Name missing or invalid. The file must be recreated
101	Error	<SoftwareCode> (800006) must be present	Software Code is missing
110	Error	<Year> (100266) must be present and = 2012	Year is missing or invalid

### 3.2 Establishment Level Validation

Seq Number	Error/Query	Validation Check	Error Message
301	Error	<LEA> (200001) must be provided and must be a valid value	Local Authority number missing or invalid
302	Error	<Estab> (200002) must be present with a valid value	Establishment No is missing

### 3.3 Pupil Level Validation

Seq Number	Error/Query	Validation Check	Error Message
<b>Pupil Identifiers</b>			
1600	Error	<DOB> (100007) must be present	Pupil's date of birth is missing
1601Q	Query	<DOB> (100007) should be between 1/9/2005 and 31/8/2006	Pupil's Date of Birth is outside expected date range
1610	Error	<Gender> (100008) must be present and a valid value	Pupil Gender missing or invalid
1500	Error	<UPN> (100001) (Unique Pupil Number) must be present except where <Estab> (200002) PRUs	UPN is missing

Seq Number	Error/ Query	Validation Check	Error Message
1510	Error	<p>If UPN is present then &lt;UPN&gt; (100001) must contain the correct check letter To calculate the check letter:</p> <p>1. Multiply the individual digits by their weights as follows: digit 2 by weight 2; digit 3 by weight 3; digit 4 by weight 4; digit 5 by weight 5; digit 6 by weight 6; digit 7 by weight 7; digit 8 by weight 8; digit 9 by weight 9; digit 10 by weight 10; digit 11 by weight 11; digit 12 by weight 12; digit 13 by weight 13.</p> <p>2. Sum the individual results, divide the total by 23, and take the remainder.</p> <p>3. Calculate the check letter from the result as follows: 0 = A; 1 = B; 2 = C; 3 = D; 4 = E; 5 = F; 6 = G; 7 = H; 8 = J; 9 = K; 10 = L; 11 = M; 12 = N; 13 = P; 14 = Q; 15 = R; 16 = T; 17 = U; 18 = V; 19 = W; 20 = X; 21 = Y; 22 = Z.</p> <p>For calculating the check letter (see above) any alphabetical character at digit 13 is accorded the same numerical value as listed in paragraph 3 above, i.e. A=0, B=1, C=2 etc.</p>	UPN invalid (wrong check letter at character 1)
1520	Error	Each pupil <UPN> (100001) must be unique across all pupils in the Estab	More than one pupil record with the same UPN
1530	Error	If <UPN> (100001) is present Characters 2-4 of <UPN> (100001) must be a valid post April 1999 LA code or a recognised "pseudo LA" code (001-005, 201-213, 301-320, 330-336, 340-344, 350-359, 370-373, 380-384, 390-394, 420, 660-681, 701-708, 800-803, 805-808, 810-813, 815, 816, 820-823, 825, 826, 830, 831, 835-837, 840, 841, 845, 846, 850-852, 855-857, 860, 861, 865-896, 908, 909, 916, 919, 921, 925, 926, 928, 929, 931, 933, 935-38)	UPN invalid (characters 2-4 not a recognised LA code)
1540	Error	If <UPN> (100001) is present Characters 5-12 of <UPN> (100001) must be numeric	UPN invalid (characters 5-12 not all numeric)
1550	Error	If <UPN> (100001) is present Characters 13 of <UPN> (100001) must be numeric or A-Z omitting I, O and S	UPN invalid (character 13 not a recognised value)
1580	Error	<Surname> (100003) must be present	Pupil with surname missing
1590	Error	<Forename> (100004) must be present	Pupil with forename missing
113	Error	Each <Assessment> should only appear once for each <Pupil> for the same combination of <Subject><Component><Method><ResultQualifier>	Assessments are either missing, duplicated or invalid
123	Error	If <ResultQualifier>=NY and <Result>=Wa or Wt then there must be a record where <ResultQualifier>=NM	Phonics Mark missing

Seq Number	Error/ Query	Validation Check	Error Message
124	Error	There must be one and only one instance where <Subject>=PHO and <Component>=CHK and <Method> =TT and <ResultQualifier>=NY	There are multiple assessments where the Subject, Component, Method? and Result Qualifier combinations are the same.
125	Error	If <Subject>=PHO, <Component>=CHK and <ResultQualifier>=NY then <Result> must be one of Wt,Wa,A,Q,D,L	The outcome result is invalid and should be one of Wt, Wa, A, Q, D or L
126	Error	There must be one and only one instance where <Subject>=PHO and <Component>=CHK and <Method> =TT and <ResultQualifier>=NM	There are multiple assessments where the Subject, Component, Method? and Result Qualifier combinations are the same.
127	Error	If <Subject>=PHO, <Component>=CHK and <ResultQualifier>=NM then <Result> must be an integer in the range 0 – 40 inclusive	The outcome result is invalid and must be in the range 0 to 40
128	Error	If <ResultQualifier>=NM and <Result>=0 - 40 then there must be a record where <ResultQualifier>=NY	Phonics outcome missing
129	Error	If <ResultQualifier>=NM and it's result is present AND there is a record where <ResultQualifier>=NY then <Result> must not be A,Q,D or L	Phonics outcome and mark invalid
130	Error	Each <Assessment> must contain <Subject>=PHO	There is an unexpected entry in the Subject container
131	Error	Each <Assessment> must contain <Component>=CHK	There is an unexpected entry in the Component container
132	Error	Each <Assessment> must contain <Method>=TT	There is an unexpected entry in the Method container

**Notes:**



## 4. Resubmission Guidance

When resubmissions are required the Serial No in the Header **must** be incremented by the provider's software. We have also assumed that the whole file will be resubmitted. Note that a resubmission will **delete and replace all the data** contained in the previous submission.

## 5. Valid values

Valid item values must be in the list or range specified in CBDS, the Common Basic Data Set, URL :  
(<http://www.education.gov.uk/schools/adminandfinance/schooladmin/ims/datamanagement/a0058744/common-basic-data-set-cbds-database>) subject to the exceptions and restrictions below.

### 5.1 LA codes

LA codes (CBDS number 200001, codeset D00004) for this collection only covers ENGLAND LAs. Coverage also excludes the following LA codes:  
001, 003, 167  
codes starting with 6 and 7 except 702 (Service Children's Education, which is valid)  
XXX, MMM and NAA

### 5.2 Establishment Types

These are the following;

Maintained schools with a Year 1 cohort  
Academies (including free schools)  
Service Children's Education Schools