The ROYAL COLLEGE of OPHTHALMOLOGISTS

Process Document

Dataset Guidelines

Authors: The Informatics and Audit Sub-Committee

Version: 2.0

Amendment Record

Issue	Description	Author	Date
0.8	First Draft	G W Aylward	24 Nov 2011
1.0	First Working Version	G W Aylward	16 Aug 2013
1.1	Draft Revision	John Sparrow	1 Aug 2018
2.0	Revision	John Sparrow	1 Oct 2018

October 2018

18 Stephenson Way, London, NW1 2HD T. 020 7935 0702
contact@rcophth.ac.uk rcophth.ac.uk @RCOphth

Contents

Section		page
1	Introduction	3
2	Structure of a Dataset	3
3	Principles	4
4	Datatypes	5
5	Process	5
6	Standard Items	7

Date of review: October 2021

1 Introduction

A dataset comprises a set of defined variables representing clinical information about a patient with a given condition. A formal RCOphth interest in ophthalmic datasets began in 2002 with the initiation of work on the cataract national dataset. Under the umbrella of the NSF for diabetes, a diabetic retinopathy screening dataset was developed and subsequently the Do Once And Share (DOAS) program supported further dataset work on cataract, glaucoma and diabetic eye care. A number of RCOphth datasets have subsequently been developed to set data collection standards in ophthalmic care and for developers of EMRs. The widespread use of the cataract national dataset has facilitated useful national audits and the development of the National Ophthalmology Database (NOD) project. Datasets facilitate the collection of information collected by different clinicians, using different systems in different locations. They are structured, easily transferable, and can form the basis of large, analyzable databases for audit and outcomes research. Datasets vary in comprehensiveness and complexity depending on the topic and intended use. A complete dataset will include all the data items required to care for a patient with a particular condition, such a dataset would be of most use to an EMR provider. An audit dataset on the other hand may include only those items which are needed to perform the audit. Ideally all ophthalmic datasets should be compatible with common data items defined in the same way. A library of data 'archetypes' for commonly used data items (e.g. visual acuity) can facilitate harmonization across datasets. This document describes guidelines for establishing new datasets.

2 Structure of a Dataset

The purpose of a dataset is to represent an agreed set of clinical information which can be collected on patients with a particular diagnosis. As well as defining the items to be collected, the dataset should describe the data type for each clinical variable. Each item should be categorized according to the scheme in the following table. Each category is a subset of the category below it:

Category	
Mandatory	Data items which are essential for all applications, and must be collected
Desirable	Advised as valuable for audit or knowledge extraction purposes
Optional	Data items which are required for some applications, and may be collected

3 Principles

A dataset should be designed to comply with the following principles;

1. Dataset Topics

Priority should be given to clinical topic areas which are high volume and / or high risk clinically. Prior to embarking on development of a dataset from scratch it is important to search out datasets which may already exist to avoid duplication of effort. Potentially useful resources include The Royal College of Ophthalmologists and the International Consortium for Health Outcomes Measurement (ICHOM)

2. The dataset should comprise information routinely collected

The intention is to not burden busy clinicians with additional work. The dataset should be constructed of items that are, or should be, recorded as part of the routine clinical management of the patient.

3. Items not required for likely analysis should be excluded unless collected as part of routine EMR use

The collection of data requires time and effort, and therefore the total number of items should be minimized where routine working does not involve EMR use. The range of analyses likely to be conducted on the data is largely predictable, and items not required for these analyses should be identified as optional.

4. Items in common with other datasets should be congruent

Several data items (for example visual acuity, IOP) will be common to many ophthalmic datasets. It makes sense that only one definition for each item is used throughout all datasets, particularly within a subspecialty.

5. The dataset should be capable of implementation in an electronic patient record

It is likely that the maximum benefit of the dataset will only be achieved when information is routinely collected using electronic patient record systems. It is therefore essential that it is capable of being implemented electronically.

6. Patient Reported Outcome Measures (PROMs)

PROMs are increasingly being recognized as an integral part of modern healthcare, including in ophthalmology. Where a Nationally or Internationally validated PROM is available a dataset should provide for its collection. A database of Clinical Outcome Assessments (COAs) can provide a helpful starting point for a search.

7. Coding of datasets

Standardizing terminology is increasingly important for medical practice, both for clinical working and management of services. The Systematized Nomenclature of Medicine -- Clinical Terms (SNOMED-CT) is the preferred NHS coding system. Ideally all datasets should be 'SNOMED-CT Coded' although currently this remains aspirational for many existing datasets.

8. The Royal College of Ophthalmologists

The RCOphth, through the Informatics and Audit Sub-committee, is keen to encourage dataset development and through the committee can provide guidance to developers. The RCOphth is able to 'Kitemark' or provide approval for datasets and place them on a RCOphth dataset register. Developers of datasets are invited to inform the RCOphth of progress.

4 Datatypes

Each item of the dataset has a data type, which in most cases should be one of the items on the following list. These correspond to data types available in database systems which underpin the storage of data within an EMR.

Type	Description
NULL	A special entity representing an uncertain or unassigned value
INTEGER	An integer value, normally unsigned (i.e. zero or positive values only)
FLOAT	A floating point value, positive or negative (avoid spurious precision)
BOOL	A value representing true or false
STRING	A value containing text (alphanumeric data) of unspecified length
ENUM	A value which represents one of a limited range of values
DATE	A value representing a date
DATETIME	A value representing a date and time
LIST	An entity containing one or more values

The term 'LIST' is not a data type but will be used to represent a 'one-to-many relationship'. This is a standard way in a database of representing data items which can vary in number (for example a patient could have one, two or more symptoms)

5 Process

The process for construction of a dataset includes the following steps:

1. Identify the need

The need for a new dataset might come from a wide range of sources, including the RCOphth, individual Ophthalmologists, researchers, or specialist societies. The RCOphth will keep a register of datasets that are in development, to avoid duplication of effort.

2. Establish a working group

The main business of the working group is to agree the fine detail of the items that are to be included. The decisions on data inclusion might be controversial. It is therefore of vital importance that the group is seen to be representative of the potential users, but not so large as to be unmanageable. Typically, a group of four to eight members is workable. Representation should be sought from as wide a range of ophthalmic working environments as possible, reflecting the 'target audience' for the dataset, and increasing the sense of 'ownership' of the dataset. All dataset development groups should include Patient and Public Involvement.

Where they exist, appropriate specialist societies should be consulted and invited to collaborate with the establishment of the working group.

The working group will require a chair who is responsible for drafting the dataset, arranging meetings or the working group, and incorporating changes in the evolving document. It is likely that the interval between the first meeting and the production of the final draft will be of the order of six months. The final draft should be reviewed, with amendments if needed, by the Informatics and Audit Sub-Committee followed by review / approval by the Professional Standards Committee.

3. Facilitation

The working group should aim to have the first meeting face to face if possible, with subsequent meetings either by teleconference or electronic communication. A web-based document store and emails will allow efficient communication, as well as storage of current versions of documents for reference by group members.

4. Testing

Members of the working group should then test the dataset in the working environment, either using paper forms, or preferably, an electronic version of the dataset. This phase of the process may result in changes to the dataset.

5. Approval

Once the dataset is completed by the working group, it is submitted to the Informatics and Audit Sub-Committee for approval. The approval criteria should include the following:

- Does the working group include appropriate representation?
- Has the appropriate specialist society been consulted and involved?
- Do all the items in the dataset have an approved data type?
- Does the dataset utilise existing data items and types (e.g. visual acuity, intraocular pressure etc.)?
- Has it been adequately tested?

6. Publication

The dataset is then published on the RCOphth website so that interested parties can download it and consider implementing it within electronic patient records, audit programs and other systems.

7. Enhancements

The use of the dataset in real systems may well reveal the need for enhancements or changes. While not needing to meet regularly, the working group should remain constituted to deal with such enhancement requests.

6 Standard Items

Certain items in datasets will be common to others (e.g. visual acuity). A parallel work stream within the RCOphth aims to produce a library of such standard items to ease construction of future datasets and facilitate data sharing. An example is patient demographics, which should include the items in the following table;

Items	Description	Values/format
Patient ID	An identifier which will uniquely identify the patient. In England and Wales this could be the NHS number. This would be removed in anonymized datasets, or replaced with a Universal Unique Identifier (UUID)	INTEGER
Age	The age of the patient in years at the time of the clinical episode. Age provides sufficient information for analysis, without also being patient identifiable data (PID), unlike date of birth. Common practice is to 'perturb' the age by +/- 3 months to avoid identification but in certain contexts (e.g. neonatology) this may be inappropriate.	INTEGER
Sex	The patient's gender	ENUM (Male, Female)
Postcode	The postcode district (outward code). This is the first part of a postcode, and generally corresponds to a post town. It gives useful information for demographic and socioeconomic analysis, without being PID	STRING
Consultant	Identifier for consultant in charge of the patient's care (separately from individuals delivering care, e.g. performing a procedure).	INTEGER

Items	Description	Values/format
Ethnic category	The ethnicity of the patient using the classification used for the 2001 census[3]	ENUM (British, Irish, Any other White background, White and Black Caribbean, White and Black African, White and Asian, Any other mixed background, Indian, Pakistani, Bangladeshi, Any other Asian background, Caribbean, African, Any other Black background, Chinese, Any other ethnic group, Not stated)
Route of referral	Route by which patient arrived in the ophthalmic department, based on who made the initial diagnosis (e.g. if an Optometrist sends a patient via the GP with a suspected diagnosis of RRD, this item would have a value of 'Optometrist')	ENUM (Optometrist, GP, Ophthalmologist from other Trust, Ophthalmologist from same Trust, General A&E, Ophthalmic A&E, New diagnosis in clinic, Other)

All datasets should adhere to the standards for the clinical structure and content of patient records, published June 2015.