The legacy of James Read

Tim Benson gives an edited version of the James Read Memorial Lecture that he delivered at the IHTSDO SNOMED Showcase Conference, Sydney, earlier this month.

James Read stood on the shoulders of giants, who shaped the early history of GP computing in UK.

In the 1960s, Dr John Preece was the first to use a computer in the consulting room. This work led directly to the Exeter GP project, and the national availability of the FP10 (comp) prescription form.

At about the same time, Dr John Perry became head of the Oxford GP record linkage project, ran a series of pioneering conferences and wrote the OXMIS codes, based on ICD-8.

And Dr Clifford Kay, who led the pioneering RCGP Oral Contraceptive Study, became the first chair of the RCGP Computer Working Party, which was established in 1978 and ran the GP-INFO-80 conference and issued a key report on Computing for GPs.

I was one of the fortunate ones to attend GP-INFO-80. Then and there, I resolved to set up Abies Informatics Ltd andto design a GP computer system, using one of the new microcomputers that had recently become available.

The original Abies GP system included an age/sex register, with 96 practice-defined on/off flags and repeat prescription printing, written in BASIC.

The next year, the system was extended to become multi-user using a version of UNIX and was renamed Abies 2.

James Read and David Markwell became our third and fourth customers respectively, and in 1982 Read was instrumental in establishingtheAbies User Group (ABUG), which he led for several years.

In June 1982 (as the result of a bit of personal ministerial lobbying at the Royal Opera House!), I obtained a grant from the government that led to city investment for a next-generation consulting room system.

In hindsight the most important choice was between staying with with UNIX and using Sculptor 4GL or using MUMPS (which was designed for medical records).

I chose to stick with UNIX and Sculptor, but this was limited to fixed length record fields. We wanted to use an existing coding scheme but none supported fixed length codes and rubrics.

The early Abies 3 systems used 4 character, practice-defined mnemonic codes (eg DIAB) with 30 character rubrics. Each line needed to fit in the 80 character width of a 80x24 screen.

xxxxxxxxx|xxxxxxxxx|xxxxxxxxx|xxxxxxxxx|xxxxxxxxx|xxxxxxxxx|xxxxxxxxx|xxxxxxxxx|

id dd/mm/yy code T rubric..... qualifier..... dd/mm/y2 AU

In December 1983, after an ABUG meeting, we decided to write our own coding scheme. This was not an extreme position. Several other projects wrote their own coding schemes at that period.

I was familiar with ICD-9, having used it while doing research on measuring patient outcomes at the Charing Cross Hospital, and I happened to have a copy in the office.

So we took ICD-9 as the starting point, to use only codes which are regularly found in general practice and to give each a short 30 character rubric.

A back of envelope calculation suggested that if we wrote 30 codes an hour, it would take 200 hours to create 6,000 codes. James Read undertook to edit the codes, and estimated that the task would take about three months. We did not account for James' obsessiveness.

James' method of working was to write the codes, rubrics and synonyms (for lookup) on a large sheet of paper using a fountain pen, and post these to Sue Uphill, our secretary in London, who entered them into the computer. I do not remember James ever using a computer himself for this or any other task.

James soon recognised that the scheme needed to cover much more than diagnoses and should cover anything that might be entered into a patient's computerised record.

No suitable coding schemes could be found for large parts of patient history, examination findings, preventive care, administrative procedures, laboratory and imaging findings and so on, so he

developed specific coding schemes for all these areas. He devoted thousands of hours to this work.

To reflect this effort and contribution, we changed the name of the codes from the Abies Medical Dictionary to the eponymous Read Codes, which were eventually delivered more than two years later than expected in mid 1986.

During this period the Abies software was also greatly improved mainly through the efforts of David Markwell and evolved to become Abies 4, which had a user interface of unsurpassed speed and elegance.

The first publication was in the British Journal of Healthcare Computing in May 1986.[i] The number of codes in the original version 1 (May 1986) were:

	Coded Terms	Synonyms	Total Terms	%
Diseases	2,598	2,575	5,173	22%
Procedures	6,023	2,483	8,506	36%
Occupations	1,749	936	2,685	11%
History	1,299	890	2,189	9%
Examination	1,480	890	2,370	10%
Prevention	1,279	460	1,739	7%
Administration	696	416	1,112	5%
Total	15,124	8,650	23,774	

Hierarchical Codes

The structure of the hierarchical classification is mapped directly by codes. In the same way as a map grid reference specifies a position on a map, each code specifies its position within the classification.

The Read Clinical Classification has four-digit alpha-numeric codes using the numerals 0-9 and the letters A-Z. The first character relates to level 1, the second to level 2 and so on. Consider code B136; this is broken down as follows:

First level	B...	Neoplasm
Second level	B1..	Malignant neoplasm
Third level	Bl3.	Carcinoma stomach
Fourth level	B136	Ca. greater curvature-stomach.

The four-digit codes increase in detail from left to right. Facilities have been built in to allow for extension within the basic framework. The alpha-numeric coding system using four digit codes allows 1,679,6l6 possible entries.

The scheme was later extended to allow lower-case letters (a-z), with the exception of a couple of letters such as O and l, which can easily be confused, giving 60 options at each level, total 604 (about 12 million options).

Automatic encoding

The classification incorporates automatic encoding. Entry of the first few letters of any term displays a list beginning with those letters from which the user chooses by line number. Consider the term `rubella'. Entry of the letters 'rub' triggers the following list:

0	H/O: rubella	1418
1	Rubella	A47.
2	Rubella + pregnancy	K2A3
3	Rubella-congenital	O25l
4	Rubella health educ.	6794
5	Rubella antibody titre	439.
6	Rubella contact	65P5

7	Rubella damage-preg.	K364
8	Rubella screen	62J.
9	Rubella vaccination	65P.

Compatibility

National and international medical classifications have been developed to facilitate the production of statistics for epidemiology and research. These systems were developed for manual recording and have been adapted latterly for use with computers.

None of these classifications covers the whole field of medicine, and none is suitable for clinical use as their coded content is not sufficiently specific.

Read aimed to be comprehensive in both breadth of cover and also in the detail of the terms used in general practice; based on existing classifications where possible.

However, large areas of medicine had not been classified before and Read extended the areas covered by the above systems to include history, symptoms, examination findings, prevention and administration (and medication).

Diseases

At the time, the International Classification of Diseases Ninth Revision (ICD-9) was the standard statistical classification of diseases, used by hospitals throughout the world.

Sections of the Read Clinical Classification which deal with diagnoses, injuries and death are directly based on ICD-9.

The Read first-digit codes A to Q correspond directly to ICD chapters, with the exception of chapter XVI (symptoms, signs and ill-defined conditions) which is covered in greater detail elsewhere. Each Read category is precisely cross-referenced to ICD-9.

This section of the Read Classification has 17 first level codes, 115 two-digit codes, 728 three-digit codes, 2598 four-digit codes and 2575 synonyms. The level of detail at each level is illustrated by the following example:

level 1	G...	Circulatory system diseases
level 2	G7..	Cerebrovascular disease
level 3	G7l.	Cerebral haemorrhage
level 4	G711	Subarachnoid haemorrhage
	G712	Intracerebral haemorrhage
	G713	Extradural haemorrhage
	G714	Subdural haemorrhage

Procedures

The International Classification of Procedures in Medicine (ICPM) complemented ICD-9 as a standard classification of surgical, diagnostic and therapeutic procedures. The Read Clinical Classification covers the whole of ICPM with the exception of the section on drugs, medicaments and biological agents.

In many cases the content and detail has been expanded to provide clinically specific rubrics. For example the results of laboratory procedures are classified as in:

level 1	4...	Laboratory procedures
level 2	46..	Urine examination
level 3	466.	Urine test for glucose
level 4	4661	Urine glucose test not done
	4662	Urine glucose test negative
	4663	Urine glucose test=trace
	4664	Urine glucose test=+
	4665	Urine glucose test=++
	4666	Urine glucose test=+++
	4667	Urine glucose test=++++

The decision to include both the procedure (urine test for glucose) and the finding (Urine glucose test negative) in the same structure was probably a mistake which has created problems ever since.

A change, made shortly after the publication of this paper, was to start sub-lists at 0 rather than 1. The lists shown here are those in the original paper, not those widely implemented.

Similarly in operative procedures mastectomy for example, is classified as:

level 1	7...	Operative procedures
level 2	7F..	Breast operations
level 3	7Fl.	Mastectomy
level 4	7F11	Breast lump local excision
	7FI2	Partial mastectomy
	7Fl3	Simple mastectomy
	7F14	Extended simple mastectomy
	7Fl5	Radical mastectomy
	7Fl6	Extended radical mastectomy
	7F17	Subcutmastect. + prosth implant
	7F18	Subcutaneous mastectomy

Sections of the Read Clinical Classification covering diagnostic procedures (including laboratory and X-ray) and therapeutic procedures (including surgery) comprise 6023 code categories and 2483 synonyms.

History/symptoms

The history and symptoms section of the Read Classification contains family, social and medical history as well as presenting symptoms.

The relevant section in ICD-9 (Chapter XVI symptoms, signs and ill-defined conditions) is incomplete and reclassification was needed to provide adequate clinical detail.

Where any history/symptom factor has gradable variables each option is offered as a separate fourth level category. For example:

level 1	1...	History/symptom
level 2	13..	Social/personal history
level 3	137.	Tobacco consumption
level 4	1371	Complete non-smoker
	1372	Trivial smoker
	1373	Light smoker
	1374	Moderate smoker
	1375	Heavy smoker
	1376	Very heavy smoker
	1377	Ex-smoker
	1378	Tobacco consumption unknown

Each term was defined. 'Heavy smoker' is 12-24 cigarettes a day or 80-160 per week, and 20 cigarettes is equivalent to 2 large cigars, 5 medium cigars, 10 small cigars or an ounce of tobacco. However, these definitions were not displayed to the user.

The history/symptoms section had 1,299 codes and 901 synonyms. History data is of cardinal importance in diagnosis and the prevention of disease and disability: "Listen to the patient, he is trying to tell you the diagnosis."

Occupations

The OPCS Classification of Occupations was the basis of this section of the Read Clinical Classification with 1749 coded occupational categories and 936 synonyms. Occupation is an important part of any patient database used for prevention or epidemiology.

Examination/signs

The classification of examination findings and signs is organised by systems. This part of the Read Clinical Classification has been classified from scratch in the absence of any other recognised classification covering patient examination.

This section comprises 19 second level, 282 third level and 1480 fourth level categories with 890 synonyms. For example, retinal examination is classified as:

level 1	2...	Examination/signs

level 2	2B..	Central nervous system exam.
level 3	2BB.	O/E – retinal inspection
level 4	2BB1	O/E – retina normal
	2BB2	O/E - retinal vessel narrowing
	2BB3	O/E – retinal A-V nipping
	2BB4	O/E – retinal microaneurisms
	2BB5	O/E – retinal haemorrhages
	2BB6	O/E – retinal exudates
	2BB7	O/E – retinal vascular prolif.
	2BB8	O/E – vitreous haemorrhages

Prevention

Preventive procedures have been classified from scratch. This is a key section of the classification particularly as computer-based prevention records and protocols could lead to major changes in the quality of patient care. This section includes:

- contraception
- pregnancy care and birth details
- child healthcare
- vaccination and immunization
- chronic disease monitoring
- health education and counselling
- screening, etc

Two examples show the level of detail provided for cervical smear screening and respiratory disease monitoring.

Cervical neoplasia screening
Cx Screen - not offered
Cx Screen - offered
Cx Screen - not wanted
Cx Screen - wanted
Cx Screen - not needed
Cx Screen - up to date
Cx Screen - not attended
Cx Screen - not reached
Cx Screen - done
Cx Screen - no result yet
Cx Screen - normal
Cx Screen - abnormal
Cx Screen + fee claimed

Respiratory disease monitoring
Initial respiratory assessment
Follow-up resp. assessment
Oral steroids last used
Intermittent drugs used more
Increasing exercise wheeze
Inhaler technique shown
Inhaler technique observed
Home nebulizer
Home oxygen concentrator
Resp. drugs side effects
Resp. treatment changed

The preventive procedures section of the Read Clinical Classification had 1279 categories and 460 synonyms.

Administration

This section covers all aspects of practice administration. Examples include the stages of patient registration and de-registration, administrative details of patient encounters, processing of claim forms, staff administration, practice finance and audit reporting.

There are 696 coded categories and 416 synonyms in the administration section of the Read Classification. For example contraception FPl001 claim status is classified as:

level 1	9...	Administration
level 2	93..	Contraception administration
level 3	93l.	FP1001 claim status
level 4	9311	FPl00l claim signed
	9312	FP1001 claim sent to FPC
	9313	FP1001 claim up to date
	9314	FP100l claim due
	9315	FP1001 claim due next visit
	9316	FP1001 claim cancelled

Drugs

A significant extension, made later in 1986, shortly after the first paper was published, was to extend the scheme to cover medicines. Drugs were given codes starting with lower-case letters a-z, corresponding to the chapters in the first edition of the BNF.

Development

One of the reasons for changing the name from the Abies Medical Dictionary to the eponymous Read Codes was to encourage other suppliers to use them too.

One of the first to take up this offer were Dr Peter Sowerby and Dr David Stables at EMIS. James set up an independent company known as CAMS (Computer Aided Medical Systems Ltd) to market the codes with royalties split equally between CAMS and Abies.

We also recognised that it would make sense for this work to be centrally supported by the Department of Health, rather than by a privately owned computer software developer.

The first major presentation was at the National Research Centre for Surgery, USSR Academy of Medical Science, Moscow in June 1986.

In 1987, the Department of Health commissioned the Joint Computing Group of the BMA's General Medical Services Committee and the Royal College of General Practitioners to evaluate clinical coding systems for use by GPs. The working party considered the following morbidity coding schemes:

1. ICD-9
2. ICHPPC-2
3. ICPC
4. OXMIS
5. Read Clinical Classification
6. RCGP classification
7. Update morbidity dictionary
8. SNOMED.

The final report (August 1988) listed the most important requirements to be:

1. Comprehensive in breadth and depth
2. Appropriate for GP usage
3. Provision for central maintenance
4. Amenable to statistical analysis
5. Compatibility with ICD-9
6. A hierarchical structure (second level requirement)
7. Accessibility of coding structure to the user (third level requirement).

The working group recommended the Read codes, with some qualifications:

1. Longer rubrics were needed for operations
2. Align to national coding schemes (ICD-9, OPCS-4, PPA Drug Index, SOC (standard occupational classification)
3. A fully resourced UK standing professional committee should be established to maintain and control the classification
4. Guidance should be provided on usage.

The Department of Health set out to implement these recommendations and after almost two years of tortuous negotiations purchased the Read Codes for £1.25m in April 1990, leading to the establishment of the NHS Centre for Coding and Classification.[ii]

Why they were so successful

Features of the first generation Read codes that made them successful were:

1. Single author
2. Fit for purpose - written by a GP for GPs
3. Comprehensive (examination findings, history, administration etc)
4. Modest evolutionary step (built on ICD-9 etc)
5. Easy to implement in software and on screen
6. Understandable by users.

The Read Codes improved on earlier classification and coding systems in several respects:

1. They were designed specifically for use by GPs in their surgery, not for epidemiology and international comparisons

2. The simple position-dependent uni-dimensional hierarchy was easy to understand by users
3. No paper version was ever published, facilitating multiple updates and extensions
4. Easy to implement in software.

Problems

No system is without problems. It is easy to make a mistake when entering data, which seriously impacted data quality.

For instance, entry of the term *physio* will give a list of options, the first being the occupation [03J1. physiotherapist]. It was easy to choose an occupation when what should have been chosen was [8H77. refer to physiotherapist]. This sort of systemic misuse is not good for data quality.

The Read Codes combine the features of a classification and a coding scheme. However, no hierarchical coding scheme can ever be multipurpose, because it will be built around a single hierarchical axis and each code is classified in one way only.

The Read Codes proved highly successful in general practice, for which they were designed. However, attempts to use the original versions in hospitals proved impracticable, primarily because the simple hierarchical scheme could reflect only one view, namely the general practice perspective.

Hospital doctors did not understand why information retrieval in one dimension was easy, but in another dimension was difficult and slow.

Position-dependent coding schemes cannot be updated. Once a concept has been placed in the classification, it is not practicable to move it, even if it has been placed in a location that is later regarded as wrong. It is not possible to add in new codes in the middle of a sequence.

Another problem is the inherent multi-dimensionality of medicine. For example, tuberculosis meningitis is a type of tuberculosis, which is an infectious disease (and is given code A130.), but it is also an inflammatory disease of the central nervous system and has another code F004.. Having two separate codes creates code redundancy, which can cause inaccuracies in hierarchy-based analysis of clinical data stored using the codes.

Being restricted to only four levels (later extended to five levels) in the hierarchy causes another problem. Consider the following hierarchy:

7.... Operations, procedures, sites
71... Endocrine system and breast operations
713.. Breast operations
7130. Total mastectomy operations
71304 Subcutaneous mastectomy

It is not possible to add a more detailed variant of this operation, such as subcutaneous mastectomy for gynaecomastia in the appropriate position because there is no 6th level.

A possible solution is to add it as a sibling alongside subcutaneous mastectomy in the 5th level with a code such as 71307. However, this creates the danger that when retrieving cases of subcutaneous mastectomy (71304), those recorded using 71307 would be missed.

The NHS Clinical Terms project was started in 1992, as a major attempt to address all of the issues listed above. Expenditure on the Read Codes between 1990 and 1998 was £32m.[iii]

The resulting scheme, which is known as Clinical Terms Version 3 (CTV3), was merged with the College of American Pathologist's SNOMED RT during 1999-2002 to create SNOMED CT.

James Read

Despite these drawbacks, and the further development of coding, there are three main reasons why the Read Codes matter today.

- Read Codes are one of the two direct predecessors of SNOMED CT. Without James Read there would be no SNOMED CT as we now know it.
- The Read Codes have been used successfully by all GPs in the UK and New Zealand and have been for the past 15 years. New Zealand still uses the original 4-byte codes, described here. They are clearly fit for purpose.
- Health management relies on comparable data. It would be unimaginable that the UK government wouldentrust 80% of the health care budget in England without GPs having coded data to monitor what is going on.

James Read was a perfectionist. He strove for excellence in everything he did. Speaking less than ten days after Steve Jobs died, I think that both men shared many characteristics. They both strove for excellence in everything they did. They both changed the world they lived in.

One of James' favourite quotes comes from George Bernard Shaw: "The reasonable man adapts himself to the world. The unreasonable one persists in trying to adapt the world to himself. Therefore all progress in the world depends on the unreasonable man."

James Read's legacy is not just his codes. He changed the way that medicine is practiced now and in the future.

---

[i] Read J, Benson T. Comprehensive Coding. *British Journal of Healthcare Computing* 3 (2) May 1986, 22-25.

[ii] Chisholm J. The Read Clinical Classification. *BMJ* 1990;300:1092.

[iii] *The Purchase of the Read Codes and the Management of the NHS Centre for Coding and Classification Select Committee On Public Accounts* Sixty-Second Report 1998www.parliament.uk