In this issue

In this issue: Ontologies a key concept in informatics and key for open definition of cases, exposures and outcome measures

Simon de Lusignan

Professor of Primary Care and Clinical Informatics, University of Surrey, Guildford, UK
Editor-in-Chief, Journal of Innovation in Health Informatics

## ONTOLOGIES A KEY CONCEPT IN INFORMATICS

Ontologies are a key concept in informatics, and the leading article *in this issue* addresses their importance.[1] Ontologies describe key concepts within a domain and their relationships. This leading article describes how to use an ontological approach to identify data sources and combine data.

We advocate that the approach to developing datasets and coding lists should also be ontological.[2] This assertion is based on a realist review of the literature[3] and an exploration of how this approach might lead to a more explicitly defined datasets when using routine data for chronic disease management,[4] integrated care[5,6] and vaccine benefit–risk research.[7]

Creating an ontology should be an explicit process so that it is clear how a case, an intervention or exposure, or an outcome measure is derived from routine data. We are adding papers describing ontologies to the type of paper we will accept in the *Journal of Innovation in Health Informatics*. Such papers should describe an ontology in the way we set out below (Figure 1) and describe the ontology and its parts.

Our recommended process for creating an ontology is to follow the three-step process shown in Figure 1. The first step is constructing the ontology *per se*; the second is to select codes relevant to the data being studied. The granularity of the ontology will need to reflect the nature of the coding and classification used in a given health care system[8] and the quality of data recording,[9] as only very rarely are all possible codes used. The final step in the process is to test if usable data can be extracted using the planned approach. If not, the ontology and coding list are revised until a usable outcome is produced. Creating a high-quality ontology is an iterative process.
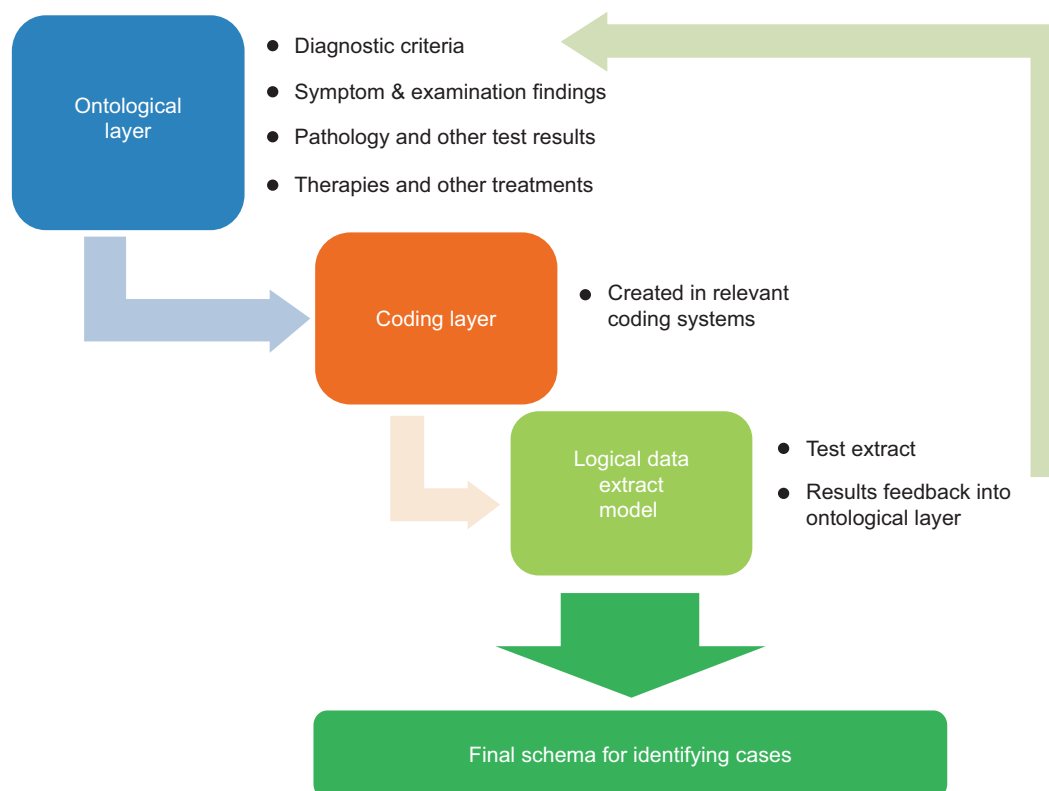
Figure 1 A three-step ontological process identifying a case from routine computer data

Step 1: Constructing the ontology

The ontological layer defines the relevant concepts. For an ontology that defines a diagnosis, this might include aetiology, diagnosis and other clinical features of the condition and its therapy. The ontology reflects the requirements and purpose of the investigation. An example of how an ontology might be created to define a case of diabetes is set out in Box 1.

Box 1 An example of how an ontological approach might improve case finding in diabetes

An ontology for diabetes would explicitly set out the criteria used in a study so that it is possible to understand how a particular prevalence might be defined. It might be restricted to one or more categories of data or require a combination (e.g. a case of Type 1 diabetes must have a Type 1 diabetes diagnostic code AND currently prescribed insulin).

Step 2: Coding layer – creating a coding list from the ontology

Each of the types of information included in the ontology should be included in the coding list. If you restrict your ontology to one or more categories of information (e.g. simply to diagnosis), then the same will apply to the coding list (in this example, it would just comprise diagnostic codes).

Step 3: Logical data extract model

The third step in using this ontological approach is to check that it is possible to extract the data you anticipate. Sometimes codes do not have sufficient granularity. Just because a code exists within a terminology, do not expect that clinicians or those involved in data entry will necessarily use it! Literature reviews, pilot searches of data sources and speaking to practitioners in the field about their data recording all help inform if your first pass model is likely to be effective in achieving its goals.

In summary, an ontological process should enable code lists used in research based on routine data to be constructed in a logical and open way. This process will enable others to use the ontology and as is, update or modify it, or apply it to other coding systems.

PAPERS IN THIS ISSUE INCORPORATING ONTLOGICAL THINKING

The final paper in this issue describes how the architecture of the computerised medical record (CMR) system can affect the prevalence of diabetes.[10] It provides a good example of why Step 3 – logical data extract model – is needed. In the UK, some CMR systems are strictly problem orientated – meaning that consultations are strictly linked to a small set number or existing problems; others allow much more flexibility of coding so that there are multiple near-synonyms for codes. This paper demonstrates why a different logical data extract model is required for each, using diabetes as an exemplar.

Using an ontological approach is highly pertinent to a qualitative analysis of the recording of diabetes data reported in this issue.[11] Robertson et al. report how carefully coded data are likely to enhance integrated care delivery, and how neglecting code data can result in information being invisible.

An ontological approach to case definition of diabetes

- **Aetiology:** Criteria that enable the validity of case identified in a population to be validated. The prevalence of most conditions is known. For example, Type 2 diabetes is rare in people under 30 years old, more common with increasing age and in men compared with women.
- **Diagnostic criteria:** Recording of a diagnostic code for diabetes, or we might stipulate classification as either Type 1 or Type 2 diabetes (people with Type 1 diabetes mellitus require insulin for survival, whereas people with Type 2 have altered glucose metabolism and may or may not require insulin).
- **Symptom codes:** Thirst, polydipsia, polyuria, and describing weight loss might be diagnostic of diabetes. The World Health Organisation (WHO) criteria for diagnosis of diabetes include abnormal blood glucose plus symptoms of diabetes; however the latter are rarely looked for in database studies.[a]
- **Examination findings compatible with the diagnosis:** Measured weight loss and smelling ketotic might imply diabetes.
- **Pathology test criteria:**
  - Fasting or random blood test results showing a raised glucose meeting the diagnostic criteria set out by the WHO;
  - Glycated haemoglobin (HbA1c) levels compatible with diabetes
  - Urine tests positive for glucose
- **Medication and prescriptions**: There are some medication and other prescribed items that imply a diagnosis of diabetes; others make the diagnosis unlikely. Some medicines, such as insulin, and some injectable and oral anti-diabetes drugs are used only in diabetes, whereas metformin is a medicine generally prescribed in diabetes but also used in other conditions. Prescriptions for testing for blood or urinary glucose or ketones make a diagnosis of diabetes more likely but not definite. For example, they may be prescribed in pregnancy or where there is impaired fasting glucose.[b]
- **Treatment or procedure codes**: There are very rare operations or other procedure codes related to diabetes. Surgery for very rare tumours – glucagonoma and phaeochromocytoma – can cure diabetes. However, this heading is included for ontological completeness.
- **Process of care codes:** There are a number of codes associated with the process of delivery of care, remuneration and administration of care which imply but do not make the diagnosis certain. There are in many ways the most complex areas of an ontology as likely to be health system specific.[c] Examples of delivery of care codes include: *'Seen in diabetes clinic'* and *'Attending diabetes clinic'.* Most people with these codes in their records will have diabetes, but some people with gestational diabetes or impaired fasting glucose may also attend. A code, in the UK, related to remuneration would include: *Excepted from diabetes quality indicators: informed dissent* – this code would be applied when someone with diabetes declines to attend for review. Its use removes them from the practice pay-for-performance target payment. Finally, *DNA – Did not attend diabetic clinic* – is an example of an administrative code.

Further information

a. World Health Organisation (WHO). *Diabetes programme*. Available from: www.who.int/diabetes/action_online/basics/en/
b. Bagheri A, Sadek A, Chan T, Khunti K and de Lusignan S. Using surrogate markers in primary electronic patient record systems to confirm or refute the diagnosis of diabetes. *Informatics in Primary Care* 2009;17(2):121–9.
c. Stone MA, Camosso-Stefinovic J, Wilkinson J, de Lusignan S, Hattersley AT and Khunti K. Incorrect and incomplete coding and classification of diabetes: a systematic review. *Diabetic Medicine* 2010 May;27(5):491–7. doi: 10.1111/j.1464-5491.2009.02920.x.

Although they do not mention the use of ontologies, this paper implies that there may be a set of codes that inform best about diabetes management.

USABILITY – A LONG NEGLECTED THEME IN INFORMATICS

We have previously asserted that usability is a long neglected theme in informatics[12] – and we welcome the paper by Joshi et al. not only for its subject matter (use of a bilingual touch-screen to provide breastfeeding education) but also for its use of a classic approach originally described by Neilson some two decades ago.[13] Neilson described the application of heuristics. Heuristics are 'practical wisdom' – an approach to solving problems – something discussed by Aristotle many centuries ago. Many of our computerised systems might benefit from the application of Neilson's heuristics!

HEALTH INFORMATION EXCHANGE AND SYSTEMATIC REVIEW PROTOCOL

The next paper is a systematic review protocol setting out how research might inform what types of enablers and blockers exist to health information exchange in low- and middle-income settings.[14] A survey of primary care providers suggests that health information exchange can support patient care, particularly when it enables key information such as medication data to be available.[15]

DIFFERENTIATING SIGNAL FROM NOISE, THE AUTOMATED MEASUREMENT OF THE RATE OF DECLINE IN RENAL FUNCTION

Ensuring we differentiate signal from noise is not just important in defining cases of diabetes.[16] Whilst inevitably renal function declines with age, there is a lot of noise[17] – particularly because measures of renal function are based on creatinine, which in turn varies depending on dietary intake of protein, muscle mass and other factors affecting protein metabolism. We publish a paper building on an approach that visualised this fluctuation that enables automated detection in fluctuation contributed by change in laboratory assay. Creatinine assays have only relatively recently been standardised. Just as the nature of the CMR system can affect how diagnostic data are recorded, difference in laboratory assay of creatinine is another, and perhaps unexpected, contributor to the difficulty in differentiating signal from noise when looking to measure the rate of decline in renal function.[18]

REFERENCES

1. Liyanage H, Krause P and de Lusignan S. Using ontologies to improve semantic interoperability in health data. *Journal of Innovation in Health Informatics* 2015;22(2):309–15. http://dx.doi.org/10.14236/jhi.v22i2.159.

2. de Lusignan S, Liaw ST, Michalakidis G and Jones S. Defining datasets and creating data dictionaries for quality improvement and research in chronic disease using routinely collected data: an ontology-driven approach. *Informatics in Primary Care* 2011;19(3):127–34. PMid:22688221.

3. Liaw ST, Rahimi A, Ray P, Taggart J, Dennis S, de Lusignan S et al. Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. *International Journal of Medical Informatics* 2013;82(1):10–24. doi: 10.1016/j.ijmedinf.2012.10.001. http://dx.doi.org/10.1016/j.ijmedinf.2012.10.001.

4. Liyanage H, Liaw ST, Kuziemsky C and de Lusignan S. Ontologies to improve chronic disease management research and quality improvement studies – a conceptual framework. *Studies in Health Technology and Information* 2013;192:180–4. PMid:23920540.

5. Liyanage H, Liaw ST, Kuziemsky C, Terry AL, Jones S, Soler JK et al. The evidence-base for using ontologies and semantic integration methodologies to support integrated chronic disease management in primary and ambulatory care: realist review. Contribution of the IMIA Primary Health Care Informatics WG. *Yearbook of Medical Informatics* 2013;8(1):147–54. PMid:23974562.

6. Liaw ST, Taggart J, Yu H, de Lusignan S, Kuziemsky C and Hayen A. Integrating electronic health record information to support integrated care: practical application of ontologies to improve the accuracy of diabetes disease registers. *Journal of Biomedical Informatics* 2014;52:364–72. doi: 10.1016/j.jbi.2014.07.016. http://dx.doi.org/10.1016/j.jbi.2014.07.016.

7. Liyanage H and de Lusignan S. Ontologies to capture adverse events following immunisation (AEFI) from real world health data. *Studies in Health Technology and Information* 2014;197:15–9. PMid:24743070.

8. de Lusignan S. Codes, classifications, terminologies and nomenclatures: definition, development and application in practice. *Informatics in Primary Care* 2005;13(1):65–70. PMid:15949178.

9. de Lusignan S and van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. *Family Practice* 2006;23(2):253–63. http://dx.doi.org/10.1093/fampra/cmi106. PMid:16368704.

10. de Lusignan S. Liaw S-T, Dedman D, Khunti K, Sadek K and Jones S. An algorithm to improve diagnostic accuracy in diabetes in computerised problem orientated medical records (POMR) compared with an established algorithm developed in episode orientated records (EOMR). *Journal of Innovation in Health Informatics* 2015;22(2):255–64. http://dx.doi.org/10.14236/jhi.v22i2.79.

11. Robertson ARR, Fernando B, Morrison Z, Kalra D and Sheikh A. Structuring and coding in health care records: a qualitative analysis using diabetes as a case study. *Journal of Innovation in Health Informatics* 2015;22(2):275–83. http://dx.doi.org/10.14236/jhi.v22i2.90.

12. Pearce C, Shachak A, Kushniruk A and de Lusignan S. Usability: a critical dimension for assessing the quality of clinical systems. *Informatics in Primary Care* 2009;17(4):195–8. PMid:20359396.

13. Joshi A, Perin DMP, Amadi C and Trout K. Evaluating the usability of an interactive, bi-lingual, touchscreen-enabled breastfeeding educational programme: application of Nielson's heuristics. *Journal of Innovation in Health Informatics* 2015;22(2):265–74. http://dx.doi.org/10.14236/jhi.v22i2.71.

14. Akhlaq A, Sheikh A and Pagliari C. Barriers and facilitators to health information exchange in low- and middle income country settings: a systematic review protocol. *Journal of Innovation in Health Informatics* 2015;22(2):284–92. http://dx.doi.org/10.14236/jhi.v22i2.98.

15. Cochran GL, Lander L, Morien M, Lomelin DE, Sayles H and Klepser DG. Healthcare provider perceptions of a query-based health information exchange: barriers and benefits. *Journal of Innovation in Health Informatics* 2015;22(2):302–8. http://dx.doi.org/10.14236/jhi.v22i2.135.

16. de Lusignan S, Hogg F and Hinchliffe RJ. Getting the signal to noise ratio right in the management of diabetes in primary care: time to stratify risk and focus on outcomes rather than process. *Informatics in Primary Care* 2010;18(4):219–21.

17. Poh N and de Lusignan S. Data-modelling and visualisation in chronic kidney disease (CKD): a step towards personalised medicine. *Informatics in Primary Care* 2011;19(2):57–63.

18. Poh N, McGovern A and de Lusignan S. Improving the measurement of longitudinal change in renal function: automated detection of changes in laboratory creatinine assay. *Journal of Innovation in Health Informatics* 2015;22(2):293–301.