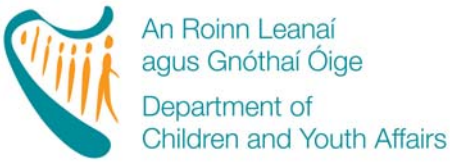

Key issues for consideration in the development of a data strategy: A review of the literature

NOVEMBER 2011

DEPARTMENT OF CHILDREN AND YOUTH AFFAIRS

The authors of this report are:

Aoife Gavin, **Colette Kelly** and **Saoirse Nic Gabhainn**, Health Promotion Research Centre, School of Health Sciences, National University of Ireland, Galway
and
Elaine O'Callaghan, Office of the Minister for Children and Youth Affairs (2008)

This research was supported by funding from The Atlantic Philanthropies

Report to be cited as:
Gavin, A., Kelly, C., Nic Gabhainn, S. and O'Callaghan, E. (2011) *Key issues for consideration in the development of a data strategy: A review of the literature*. Dublin: Department of Children and Youth Affairs. Available at: www.dcya.ie

Department of Children and Youth Affairs
43-49 Mespil Road
Dublin 4
Tel: +353 (0)1 647 3000
Fax: +353 (0)1 667 0826
E-mail: contact@dcya.gov.ie
Web: www.dcya.ie

Published by Government Publications, Dublin

Contents

Executive Summary

This report, supported by funding from The Atlantic Philanthropies and completed in 2008, highlights key issues for consideration in the development of a data strategy. The Irish Government has stipulated that each Department should develop a formal data or statistics strategy as part of its information strategy.

Chapter 1 highlights the purpose of a **comprehensive data strategy**. It is important that the value of data is understood within an organisation, both for the end users and for the production of data itself. In addition, there is a need for all stakeholders to understand the role of research and data within an organisation, and to this end three models of research utilisation are discussed – the problem-solving model, the interactive model and the dialogical model.

Chapter 2 explores the **management challenges of data systems**, with an examination of data quality issues. The RADAR model, developed by the United Nations Statistics Commission, is proposed to aid in effectively developing and improving the performance of data or statistics strategies. It is necessary to consider, agree and implement the well-managed documentation of data flow within an organisation as part of a data strategy. The definition of 'data quality' and ways in which it can be measured are also discussed . Two existing data quality frameworks are presented, one from Canada and the other from New Zealand. The 2005 *Data Quality Framework* by the Canadian Institute for Health Information (CIHI) uses 5 dimensions of data quality: accuracy, timeliness, comparability, usability and relevance. The New Zealand Ministry of Health's data quality framework is based on the CIHI model and aims to achieve consistent and accurate assessment of data quality in all national health data collections so as to enable improved decision-making and policy development in the health sector.

Chapter 3 explores issues of **data integration**, which refers to a set of processes that allows organisations to access and locate fragmented data, as well as to aid in the creation of accurate and consistent views of the issue that needs to be addressed. There are three main approaches or models of data integration – data warehousing, metadata and data brokering. Each of these have specific advantages and challenges, highlighted within the discussion. Following on from this, a discussion of the legal issues concerned with data integration focuses on the issues of consent, anonymisation, data ownership, research, electronic health governance, liability issues and other systems issues. In terms of data protection, international examples are presented to illustrate the issues with specific regard to children. Finally, ethical issues of data integration are explored, concluding with a discussion on the challenges to the development and use of electronic health records.

Chapter 4 illustrates three **examples of data strategies** from international sources. Data strategies from the Netherlands, Australia and the USA are explored to provide insight into the development of a data or statistics strategy.

This report highlights that the development and production of a data strategy is both complex and multifaceted. The cooperation of multiple key players, as well as interdepartmental collaboration, is essential. There are statistical programs and companies available to be third-party members of a data strategy. Nevertheless, the development of a data strategy for children's statistics raises many legal and ethical challenges that require detailed consideration.

1. The need for and purpose of a comprehensive data strategy

A Government decision in April 2003 stated that each Department should develop a formal data/statistics strategy as an integral part of its information strategy. In order to develop such a strategy, it is important to identify both the operating environment and data needs of a department. The Office of the Minister for Children and Youth Affairs (OMCYA; now the Department of Children and Youth Affairs) identified a need for research to be undertaken in this area. The present research was conducted jointly between the Health Promotion Research Centre, NUI Galway, and the Research Unit of the OMCYA, supported by funding from The Atlantic Philanthropies. Completed in 2008, it identifies the key issues to be considered in the development of a national data strategy.

In 2004, the National Statistics Board (NSB) published a paper entitled *Best Practice Guidelines for the development and implementation of formal data/statistics strategies in Government Departments* (NSB, 2004). In general, the guidelines outline 7 processes involved in developing a data strategy, namely: identification of data policy needs, analysis of internal data availability, identification of data help by other departments, identification of data gaps, maximising the value of data, using statistics effectively and developing data use protocols. The information included within the present report will address each of these processes.

It has been stated that '*the basic task of official statistics is to provide reliable, high-quality and relevant statistical information for citizens, general government, businesses, scientific research and international organizations*' (Jeskanen-Sundström, 2007). With rapid advances in both information and communication technologies, there has been accelerated development of society. Changes in both the social and economic domains have had two major impacts on statistical organisations, as well as on governmental departments concerned with statistics (Jeskanen-Sunström, 2007). First, the statistics that are used to help describe society must constantly change in order to reflect changes in society itself; this means there is an onus to ensure that the production of statistics remains relevant to the needs of an evolving society. Secondly, management of statistical operations changes with the shifting environment and culture of organisations; this means that management systems need to reflect the growing demand on statistical organisations.

Adelman (2001) argues that in most organisations the potential value of data is not adequately understood. Data are often considered to be the property of the department that commissioned its production and is rarely shared between departments or organisations. In addition, there are very few organisations with a clearly defined data strategy. As a result, data are often considered on a needs basis, with much overlap and unnecessary work occurring. Adelman suggests, '*Not having a data strategy is analogous to a company allowing each department and each person within each department to develop their own chart of accounts. The empowerment would allow each person in the organisation to choose his or her own numbering scheme … the resulting chaos is obvious*'.

Two basic purposes of developing a data/statistical strategy are outlined by Sundgren (1996). The first is the end-user oriented purpose, which is about providing support to potential users of statistical information through, for example, Internet data dissemination systems. A potential end-user of statistical information needs to be able to identify, locate, retrieve, process, interpret and analyse statistical data that may be relevant for a task that the user has at hand (Gillman *et al*, 1998). The second purpose is production-oriented and is concerned with the planning, design, operation, processing and evaluation of statistical surveys.

It is important to understand the statistical relationships involved in developing a data/statistical strategy. Grossmann (2004) provides an overview of these statistical relationships, in which the statistical dataset refers to a statistical population which is based on a statistical unit. It is obtained through various production methods and contains numeric information. It carries structural relationships which are defined by statistical variables.

The role of research

Harries *et al* (1999) argue that there is often tension between research and services management, and that in order to ensure that research is effectively feeding into evidence-based policy and practice, it is important that there is a shared model of research utilisation. There are three main models of research utilisation.

- **Problem-solving model** – proposes that research is used to identify the gaps in knowledge and subsequently fill those gaps (Elliot and Popay, 2000).
- **Interactive model** – one where '*research is one of several knowledge sources on which policy-makers draw in an iterative process of decision making*' (Elliot and Popay, 2000, p. 462). Within this model, researchers must try and gain a position of influence with policy-makers.
- **Dialogical model** – proposes that '*knowledge is created through, not despite, interaction*' and that research is borne of the social world.

Each of these three models provide a framework for the development of a relationship between research (and the data it produces) and evidence-based policy. Harries *et al* (1999) highlight the importance that all key players have a clear understanding of which model of research utilisation best serves their needs.

However, there are other crucial considerations that deserve examination, including management challenges and issues around data integration. These are now considered in turn. Chapter 2 considers management challenges, the issues of data documentation and data quality. Chapter 3 examines data integration and the issues of data repositories, unique identifiers and electronic health records, as well as related legal and ethical issues.

2. Management challenges of data systems

Any effective data strategy needs to deliver the ability to cope with new and ever-changing data requirements, the capability to maintain high-quality statistical products and services, and the overall ability to produce value for customers and other end-users (Agosta, 2007). Strategic management of statistical information is an ever-growing challenge. The United Nations Statistics Commission proposes the RADAR model to help effectively develop and improve the performance of statistic strategies (Jeskanen-Sundström, 2007):

- Determine the **R**esults that are aimed for.
- Plan and develop an **A**pproach.
- **D**eploy the approach in practice.
- **A**ssess and **R**eview the approach and its deployment.

In order to determine the desired results, there are a number of key issues that need to be considered as part of a management approach to data. First, we consider issues of data documentation and data quality. (Chapter 3 will consider issues of data integration.)

Documentation

Documentation is a key element in the development of any statistical database or registry (Agosta, 2007). In essence, documentation provides the proper context of information dissemination and storage. There are two main types of documentation that need to be maintained:

- **Data quality documentation** for users is given to users so that they can determine the quality of the data and its 'fitness of use'. It should include 7 elements:
    - mandate or purpose of the database;
    - the population included;
    - data elements and their conceptualisation;
    - major data limitations;
    - data collection and non-response;
    - coverage (description of the frame);
    - external comparability.
- **Methods documentation** should include detailed documentation of the methodology of the database. The purpose of this documentation is to document the data flow within a database.

Data quality

Before a discussion of data quality, it is important to note that although there has been an extensive amount of literature dedicated to data quality, there is no consensus on a good set of data quality dimensions (Wand and Wang, 1996). Data quality is a concept that is not well defined within current practice. In general, data are generally considered to be quality data when it is deemed appropriate to use for the purpose in question. All of the literature indicates that data quality is difficult to measure and that one measure of data quality is not ideal for all situations (CIHI, 2003). Klein and Rossin (1999) note that there is no single definition of data quality accepted by researchers working within the discipline of health. The following definition of data quality is proposed by Redman (2001): '*Data are of high quality if they are fit for their intended uses in operations, decision-making and planning. Data are fit for use if they are free from defects and possess desired features*.' Iezzoni (1997) indicates that it is imperative to determine who is accountable for data quality within any organisation.

Wang *et al* (1993) provide useful definitions for understanding data quality in terms of '*conformance to requirements*'. According to these authors, there are 7 key definitions that operationally define data quality:

- **A data quality parameter**: A subjective dimension by which a user evaluates data quality (i.e. source credibility, timeliness).

- **A data quality indicator**: Provides objective information about the data (i.e. source, time, and collection methods).
- **A data quality attribute**: Collective term that includes both parameters and indicators.
- **A data quality indicator value**: A measured characteristic of the stored data.
- **A data quality parameter value**: The value determines a quality parameter.
- **Data quality requirements**: The documentation of data quality measures, which allows for users to retrieve data of a specific quality.
- **The data quality administrator**: Person (or system) whose responsibility it is to ensure that the data conforms to the quality requirements.

These 7 definitions can be used to provide clarity to the complex nature of data quality. Strong *et al* (1997, p. 104) have produced a list of data quality categories and dimensions that can be related to the above definitions (*see Table 1*).

Table 1: Data quality categories and dimensions

Category	Dimension
Intrinsic	Accuracy Objectivity Believability Reputation
Accessibility	Accessibility Access security
Contextual	Relevancy Value-added Timeliness Completeness Amount of data
Representational	Interoperability Ease of understanding Concise representation Consistent representation

Source: Strong *et al* (1997)

Wang *et al* (1995, p. 623) outline the elements of a data quality framework; they begin by stating that '*for organisations to be best served by their information systems, a high degree of data quality is required, and the need to ensure this quality has been addressed by both researchers and practitioners for some time*'. Information systems can be compared to manufacturing systems where data are seen as the raw materials and information is the final product, or output. Table 2 draws an analogy between information systems and manufacturing systems.

Table 2: Information *versus* manufacturing systems

	Product manufacturing	Data manufacturing
Input	Raw materials	Raw data
Process	Materials processing	Data processing
Output	Physical products	Data products

Wang *et al* (1995) propose the use of the International Organisation for Standardisation's ISO9000 for the development of a data quality framework. The framework they present is based on the aim and objectives of the ISO9000 and contains 7 elements:

- management responsibilities;
- operation and assurance costs;
- research and development;
- production;
- distribution;
- personnel management;
- legal function.

This data quality framework by Wang *et al* (1995) takes a unique approach to data quality in a methodological and systematic way. It could also be easily amended and integrated with other data quality frameworks. To begin, senior management should be responsible for the development of a data quality policy or strategy. This data quality policy should be adhered to and changed if and when the need arises. The next element highlights the fact that unlike raw materials, data can be used repeatedly. This, however, illustrates the need for the quality of the data to be ensured to the highest possible standards – poor data can have costly effects. Within research and development, there is a need to develop or adopt an operating system that can perform the stipulations of the data quality policy. The production element of the data quality framework establishes the need for continuous checking of the data with various quality control measures in order to ensure data quality on raw data, work in progress and final data products. Distribution is concerned with the storage of data, as well as quality documentation (*see above*). Personnel management highlights the importance of ensuring that all involved with the data are cognisant of the data quality strategy. Finally, the legal function element is primarily concerned with all of the legal implications of the data itself.

Wand and Wang (1996, p. 88) illustrate the importance of a distinction between the internal and external view of information systems. They state that '*the external view is concerned with the use and effect of an information system*', while '*the internal view is concerned with the construction and operation of the information system*'. Table 3 outlines the data quality dimensions that are related to both the external and internal views of such systems.

Table 3: Internal and external perspectives on data quality

	Dimensions
Internal view (design, operation)	**Data-related** Accuracy, reliability, timeliness, completeness, currency, consistency, precision **System-related** Reliability
External view (use, value)	**Data-related** Timeliness, relevance, content, importance, sufficiency, useableness, usefulness, clarity, conciseness, freedom from bias, informativeness, level of detail, quantitativeness, scope, interpretability, understandability **System-related** Timeliness, flexibility, format, efficiency

Source: Wand and Wang (1996)

The Canadian Institute for Health Information (CIHI) published a document entitled *The CIHI Data Quality Framework* in 2005 with the aim of providing a common objective approach to assessing data quality. The framework provides a tool for database and registry areas to utilise a standard method of assessing data quality. It calls for ongoing assessment throughout the lifecycle of a piece of work. The work cycle is a three-component approach that involves planning, implementing and assessing. The *planning stage* includes activities necessary to prepare and prioritise the processes required for a database or registry, as well as the design of any changes needed. The *implementing stage* includes developing the processes needed and applying them to the database or registry (specifically, this may mean collecting data). Finally, the *assessing stage* involves evaluating the quality of the database or registry, and determining if any changes to the processes are needed.

This three-stage cycle has been designed to be flexible in that it could apply to a database or registry at any point within its development cycle. The CIHI highlight that '*It is more desirable to monitor the quality of incoming data on an ongoing basis than to wait until all data have been received to assess the quality*'. The assessment of data quality is accomplished through the use of an assessment tool.

The CIHI have defined data quality as '*fitness for use*'. This depends primarily on the use being discussed and the standards of the user. In order to establish an operational definition of data quality, 5 dimensions have been identified, namely: accuracy, timeliness, comparability, usability and relevance (*each discussed below*). Within the literature, there are no clear ways of defining data quality. However, the CIHI has the most developed assessment tool. Therefore, it will be used to describe the 5 dimensions listed above. The assessment tool is comprised of dimensions, characteristics and criteria. More specifically, each dimension is broken down into characteristics and each characteristic is made up of various criteria. Within each of the 5 dimensions, there are particular criteria that can be utilised to assess a rating for that particular element of data quality. The CIHI recommend using a system whereby each criterion is scrutinised and given a rating of either 'met', 'not met', 'unknown' or 'not applicable'.

Accuracy

The accuracy of a dimension refers to how well information in or derived from a database or registry reflects the reality it was designed to measure. The accuracy of a database depends of many factors, many of which are hard to measure. There are 3 important questions to be considered concerning data accuracy:

- Are all the appropriate data present?
- How good are the data?
- What is done with the data?

To answer these questions, the CIHI have developed 7 characteristics that outline data accuracy: coverage, capture and collection, unit non-response, item non-response, measurement error, edit and imputation, and processing and estimation.

- **Coverage** can refer to either over- or under-coverage. This occurs when there is a difference between the population of reference and the frame used.
- **Capture and collection** refers to the practices that are used with dealing with the data suppliers and during data entry.
- **Unit non-response** occurs when entire records are missing from the database. Non-response can occur at varying levels, and therefore it is vital to get some measure of the amount of data missing from a database.
- **Item non-response** occurs when a record is received and has some blank data elements that should not be blank.
- **Measurement error** of data can occur in 3 components. It is important to note that these components have the potential to overlap:
    - *measurement error* assesses to what degree the values reported match the values that should have been reported;
    - *bias* assesses to what degree the difference between the reported values and the values that should have been reported occurs in a systematic way;
    - *consistency* assesses the amount of variation that would occur if repeated measures were done.
- **Edit and imputation** needs to be assessed in order to determine the accuracy of the data. Each organisation has mechanisms in place that dictate both editing and imputation of data. Editing is the process of identifying blank or missing data, and imputation is what dictates what should be done with the missing or blank data.
- **Processing and estimation** focuses on whether programs or systems affect the data quality. There is a need for documentation for all of the data processes that occur throughout the entire procedure.

Timeliness

Timeliness refers to how current or up to date the data are at the time of release, by measuring the gap between the end of the reference period to which the data pertain and the date on which the data became available to users. However, it must be noted that within the literature there is caution that if too much emphasis is placed on timeliness, data accuracy may be compromised. The characteristics of timeliness include data currency at the time of release, as well as documentation currency. As a result, 2 key questions arise:

- Are data made available in a reasonable amount of time?
- Are key documents released on time?

Data currency is the key component of timeliness and is measured by taking the difference between the date of release and the last date to which the data relate. The CIHI (2006, p. 30) point out that '*If the methods used to process and analyse the data are as accurate and efficient as possible, the data will not be unnecessarily delayed*'.

Comparability

Comparability refers to the extent to which databases are consistent over time and use standard conventions, making them similar to other databases. Comparability facilitates the understanding, interpretation and maintenance of the data. Databases that are comparable will use the same data definitions, collect similar types of data and have the potential for record linkage with other similar databases. The characteristics of comparability include data dictionary standards, standardisation, linkage, equivalency and historical comparability. In order to determine comparability, the following questions must be addressed:

- Does the database use standard definitions for data definitions?
- Can common groupings be derived from the data?
- Can databases be joined via a common data element?
- Are data values being converted correctly?
- Are data comparable over time?

The CIHI provides guidelines on how to measure effectively each of the above dimensions.

Usability

Usability reflects the ease with which a database or registry's data may be understood and accessed. There are several factors that contribute to the usability of a database's data. Firstly, the greater the number of limitations on the data, the more difficult it is to interpret the data. In order to aid interpretation, key data users should be informed of any known limitations. The purpose of assessing usability is to identify any problems that are related to interpretability. The characteristics of usability include accessibility, documentation and interpretability. The following questions must be addressed in order to determine the usability of a database or dataset:

- How readily accessible are the data?
- How well documented are the data?
- How easy is it to understand the data?

Relevance

Finally, the relevance aspect of data quality refers to the degree to which a database or registry meets the current and potential future needs of users. In order to maintain relevance within a database, there must be communication between the key users and the stakeholders. The purpose of assessing relevance is to determine how well a database can adapt to change and whether the database is perceived as being valuable. The characteristics of relevance include adaptability and value. The following questions must be addressed:

- Can user needs be anticipated and planned for?
- How valuable are the data?

The above discussion shows that data quality is a complex construct to define. There are many ways in which various organisations and departments have approached measuring and assessing data quality. The Canadian Institute for Health Information (CIHI) provides a comprehensive and systematic approach to data quality, which could be adapted to meet the needs of the Office of the Minister for Children and Youth Affairs (now the Department of Children and Youth Affairs) in establishing a data strategy. The next section will provide an outline of existing data quality frameworks from Canada and New Zealand, which provide concrete examples of the information provided above on data quality.

Existing data quality frameworks

Canada

The first page of *The CIHI Data Quality Framework* (2005) states, '*Maintaining and enhancing data quality at CIHI is essential to ensuring that the organization can deliver on its mandate of providing timely, accurate and comparable information to inform health policies, support the effective delivery of health services and raise awareness among Canadians of the factors that contribute to good health*'. There are 6 key action areas identified by the CIHI that are to be fostered by the development of the data quality strategy. These include:

- foster a data quality culture within CIHI and in the broader health sector in general;
- strengthen CIHI's data quality infrastructure and capacity;
- cultivate the data supply chain;
- enhance external collaboration;
- establish a dedicated fund for fast track priority data quality projects;
- communicate and consult on CIHI's data quality strategies and action plans.

As outlined in the sections on 'Data quality' above, the CIHI developed an assessment tool which draws on 5 dimensions of data quality: accuracy, timeliness, comparability, usability and relevance.

New Zealand

The Health Information Steering Committee (2005) of the New Zealand Ministry of Health stated that '*Organisations are becoming more dependent on data and virtually everything the modern organisation does depends on and creates large volumes of data*'. The Ministry of Health recognised that there was a growing demand for a data quality framework to be developed as a tool for the assessment of data quality within the organisation (Kerr, 2002). For the *Health Information Strategy for New Zealand 2005*, the Ministry drew upon Carson's (2001) work on the development of a data quality framework for the International Monetary Fund. Carson outlines that an assessment tool for data quality needs to have the following characteristics:

- comprehensive coverage of the dimensions of quality and the characteristics that might represent quality;
- balance between rigor desired by an expert and the bird's eye view desired by a general data user;
- structure, but flexible enough to be applicable across a broad range of data collections;
- lead to transparent results;
- be arrived at by drawing on best practice.

Overall, the aim of the New Zealand *Health Information Strategy* is to deliver '*a data quality framework that allows for the consistent and accurate assessment of data quality in all national health data collections held by the Ministry of Health, which will enable improved decision-making and policy development in the health sector*' (Ministry of Health, 2005, p. 8). The New Zealand data quality framework is adopting and modifying the CIHI's data quality framework. After thorough evaluation, it was deemed to be the most robust and workable data quality framework readily available.

3. Data integration

A second managerial task when developing a data/statistical strategy is to adopt and implement an appropriate data integration framework. Data integration is a tool that allows organisations to access and locate fragmented data, as well as to aid in the creation of an accurate and consistent view of the issue being addressed (Kimball *et al*, 1998). Data integration is concerned with the issue of combining data from different sources and providing the user with a combined or unified view of these data. Hull (1997) highlighted the need for the development of techniques that support integrated access to databases.

There are two approaches to data integration that occur within database research (Hull, 1997). The first is an **on-demand approach**, where data are extracted from sources only when queries occur. This approach involves a two-step process:

- Query is received, the appropriate information is sourced and appropriate sub-queries are generated for each information source.
- Obtain results of queries from information sources, translate or merge results where appropriate and return the final answer to the end-user.

The alternate to this approach is the **in-advance approach** (commonly referred to as data warehousing, *see below*), which occurs in the following two-step process:

- Information from all relevant sources are translated, filtered and merged with relevant information from other sources and stored in a centralised repository.
- When a query is received, the information is sought from the repository rather than from the original information source.

Models of data repositories

There are 3 main models of data repositories – data warehousing, metadata and data brokering. Each is discussed below.

Model 1: Data warehousing

The concept of a data warehouse was first introduced in 1992 by W.H. Inmon. According to Srivastava and Chen (1999), the goal of the data warehouse is to provide a mechanism that allows organisations to use enterprise-wide information for key decision-making activities. The data warehouse is '*used to describe a subject-oriented, integrated, non-volatile and time variant collection of data, in support of management's decision-making*' (*ibid*). Overall, providing integrated access to multiple, distributed, heterogeneous databases and other information sources has become one of the leading issues in database research.

The basic architecture of a data warehousing system, as illustrated by Widom (1995), has three main components: information sources, wrapper/monitor and integrator. The foundations of the warehouse are the various *information sources* that are available on any given topic. These sources may represent any number of surveys, knowledge bases, legal documents, etc. Connected to each information source is a wrapper/monitor. The *wrapper* component of this module is responsible for translating information from the raw source into the formal data model used by the warehousing system. The *monitor* is responsible for automatically detecting changes of interest in the source data and reporting them to the *integrator*.

One of the issues arising in data warehousing is that of data quality. Ballou and Tayi (1999) explore issues of data quality in data warehouse environments. Their main concern is the use of the same data for a multitude of purposes, which can be different from the purposes of the data originally. They suggest that quality assurance measures must be carried out at each phase of warehousing – planning, implementing and maintaining. In order to do this, it is proposed that the process be approach systematically. In general, data warehouses are established to support a division of an organisation and understanding the needs of that division will allow for better data quality.

Srivastava and Chen (1999) provide a thorough explanation of the creation of a data warehouse. The initial decision in developing a data warehouse is deciding on the architectural approach to take. There appears to be 3 main architectural approaches: database conversion, database synchronisation and federated databases.

- **Database conversion** is similar to data integration. The various source databases are fed through a database conversion engine with the aim of developing an integrated target database. Database conversion is usually conducted on large volumes of data, where the databases are mapped into a global model.
- **Database synchronisation** refers to coordinating existing data warehouses periodically. The volume of data involved is considerably less than in database conversion; also there appears to be problems in the integration of the data involved in this proposed architecture.
- **Federated databases** are driven by specific queries. Integration of appropriate data occurs at the time of query and therefore the original databases remain as independent entities. This third model is the most complex and requires much more in-depth work by statisticians and has the potential to be more costly over time.

In order to create and maintain a data warehouse, organisations require adequate technical support. In examining various data strategies, it appears that Oracle RDBMS is one of the software tools used to house data. (The acronym RDBMS stands for 'relational database management system'.) This technology is utilised both by the Finnish National Research and Development Centre for Welfare and Health (www.stakes.fi) and by the Australian Bureau of Statistics (ABS) (www.abs.gov.au/).

Finally, the work of Srivastava and Chen (1999) highlights the following issues which are particularly relevant to the creation of a data warehouse:

- **Semantic issues**, revolving around the notion that any given measure represents the same or different real-world entity of any other similar measure. For example, the measures of income and salary in two separate databases.
- **Scalability**: The function of a data warehouse is to store information about the database and thus it has the potential to grow exponentially. There does not appear to be any final solutions or proposals for managing the size of the data warehouse and this must be addressed by those involved in creating and maintaining the warehouse.
- **Incremental updates**: The dynamic nature of a data warehouse lends itself to incremental updates. With the addition of new information to the warehouse, it must be carefully integrated with pre-existing information. This involves the use of incremental algorithms and indices. More research is needed in order to develop and provide solutions for this particular issue.

Example of a data warehouse – Australian Bureau of Statistics

The Australian Bureau of Statistics (ABS) is one example of an organisation that has adopted a data warehouse approach using the Oracle technology program. The following information is based on publications by Oracle (2006) for its customers and could therefore be considered as somewhat subjective. The Oracle data warehouse allows ABS to store large amounts of data and information in one single location and in effect allows for more efficient analytic capabilities. The data warehouse is known as the 'Output Data Warehouse' and it allows the organisation to store and manage data from one location, as well as generate a range of products in different formats from a single data source. The benefits of using the Oracle data warehouse (as quoted in Oracle, 2006) include:

- 'increased efficiency by re-engineering the input supply chain, enhancing data capture and storage techniques;
- enabling swift querying, performance tabulations, and data aggregation by adopting a star schema database structure;
- facilitating fast response times and providing the flexibility to access and view data using Oracle Discoverer;
- ensuring analytic computations can be run faster by pre-assembling certain elements of the results;
- protecting data from unauthorised viewing through the use of Oracle Fine Grained Access Control to define roles and assign privileges to users;

- easing database administration and improved service levels with Oracle Enterprise Manager;
- enabling implementation of a conceptual framework for metadata based on international standards.'

Model 2: Metadata

Statistical metadata is defined by the United Nations Economic Commission for Europe (UNECE) as '*descriptive information or documentation about statistical data*' (UNECE, 1995). It facilitates sharing, querying and understanding of statistical data over the lifetime of the data. Metadata is simply described as '*information about information*' (Gillman *et al*, 1998). In essence, it provides a structured set of details about a particular information source. This information is useful to those who store it as well as those who wish to access and use it. Overall, metadata makes information resources easier to manage and find. Lenz (1994) describes two types of statistical data: (1) microdata, which is data on the characteristics of units of a population (such as individuals, households or establishments) collected by a census, survey or experiment; and (2) macrodata, which is data derived from microdata by statistics on groups or aggregates (such as counts, means or frequencies).

Sumpter, cited in Gillman *et al* (1998), outlines 3 components of metadata:

- **Systems** – the information about the physical characteristics of the datasets, such as location, record layout, database schemas, media and size.
- **Applications** – the information about sample designs, questionnaires, software, variable definitions, etc.
- **Administrative** – the management information, such as budgets, costs and schedules.

Metadata repository tools are divided into several types: collection, registration, crosswalk, maintenance and query (Dao and Perry, 1996). Heerschap and Willenborg (2006) outline the advantages of using a metadata approach:

- it allows for accessibility of data;
- it applies pressure for more coordination of concepts and definitions used;
- it enhances reproducibility;
- it contributes to efficiency.

The Irish Public Service Metadata Standard describes the following benefits of metadata (IPSMS, 2007):

- the adoption of a single metadata standard across resource creators and providers facilitates precise and accurate information retrieval;
- by searching resources by a particular descriptor, for example; title or creator, the result will meet the user's needs more precisely;
- metadata can be used to provide a range of pertinent details about a resource, some of which may be missing from the body of the resource itself;
- increasing the precision of searching is more efficient in terms of user time;
- metadata also helps with the maintenance of information resource collections. It can be used to identify information that needs to be updated or archived, and individuals responsible for maintaining a resource;
- metadata can help 'join up' service providers. A standard approach to describing and storing information resources is a basis to searching and retrieving resources, which may be distributed across a number of collections in difference locations.

Adelman (2001) outlines the various components of metadata that need to be examined in order to develop a cohesive data strategy. These components include, but are not limited to, management support for metadata; deciding on which metadata to capture; determining responsibility for capturing metadata; examining how the metadata will be captured; producing tools that produce metadata; and sourcing software that will support all of this work. Without metadata, the knowledge contained within information sources cannot be effectively used.

Example of metadata – Population Health Observatory

The main concern for Ireland and Northern Ireland's Population Health Observatory (www.inispho.org/) was to promote interoperability. The INIsPHO uses a modified version of the Dublin Core Metadata Element Set for the management and use of its data. The goal of the element set is '*to foster the widespread adoption of interoperable metadata standards and promote the development of specialised metadata vocabularies for describing resources to enable more intelligent resource discovery systems*' (Kavanagh *et al*, 2005). In essence, the Dublin Core Metadata Element Set proposed 15 metadata elements, which can be adopted by various departments that need to both organise and classify their information (*see Table 4*).

Table 4: Dublin Core Metadata Element Set

Metadata element	Definition
Title	A name given to the resource
Creator	An entity responsible for making the content of the resource
Subject	A topic of the content of the resource
Description	An account of the content of the resource
Publisher	An entity responsible for making the resource available
Contributor	An entity responsible for making contributions to the content of the resource
Date	A date of an event in the life of the resource
Type	The nature or genre of the content of the resource
Format	The physical or digital manifestation of the resource
Identifier	An unambiguous reference to the resource within a given context
Source	A reference to a resource from which the present resource is derived
Language	The language of the intellectual content of the resource
Relation	A reference to the related resource
Coverage	The extent or scope of the content of the resource
Rights	Information about the rights held in and over the resource

Source: Adapted from INIsPHO (2007)

These metadata elements can be modified to suit the needs of any given organisation. They can be used as the basic format for the creation of metadata to be used in the development of a data strategy.

Model 3: Data brokering

Data or information brokering refers to a system that is capable of retrieving information about services, products or specific queries via the Internet from numerous databases for both human and computer-based individuals (Rigby *et al*, 2005). The basic architecture of a data brokering system is described by Fikes *et al* (1996).

The components of a health data broker follow a query process that flows from one process to another. To begin, a query is created, this query is processed and the various attributes are identified and authorised (Rigby *et al*, 2005). Next, the results are generated and the content is authorised, which then goes on to the results processor that determines and validates the results based on the initial query. The vision of a heath data broker is that an end-user will log onto the system and submit a query that will be processed in real time.

The information broker is a relatively new idea in healthcare management and databases. There are two issues to consider when examining a data integration brokering system: (1) organisations are committing vast amounts of information to computer systems (Fikes *et al*, 1996); and (2) information technology, particularly networking, is increasing connectivity to allow for integration of such information (Rigby *et al*, 2005). Budgen *et al* (2007, p. 34) highlight the issue that '*any mechanism that collects and stores personal information from multiple sources must consider not only the technical challenges, but also the rights of the individuals and organisations providing the data*'. As outlined below, data acquisition has centred primarily on the electronic healthcare record. However, healthcare needs require information from a spectrum of potentially independent and autonomous sources (Rigby *et al*, 2005).

The central argument towards the use of an integrated data brokering system is highlighted by Budgen *et al* (2007, p. 35): '*All too often, retrospective studies of tragedies involving serious or fatal neglect [in children] show that individuals from various agencies had some information about apparently minor incidents, but saw it in isolation. Had data on all these individual concerns be brought together at a single point in time, the true picture would have rapidly emerged and that child may have been protected*.'

Challenges to data integration

The development of a data integration strategy is to a large extent dependent on there being broad agreement on a range of other issues, particularly the interrelated concerns around national identifiers, electronic health records and data protection. Some of the issues are technical or operational, but others are legal and ethical, and are discussed below. Although the debate is well documented in relation to electronic health records and electronic patient records, the underlying principles are applicable to the whole gamut of personal and structural information relevant to individuals.

Unique individual identifiers and electronic health records

Within the sphere of healthcare, useful models of shared care information have been developed. One such model deals with an individual's electronic health record (EHR), which contains relevant medical information alongside related non-medical information. The Article 29 Data Protection Working Party[*] (2007) defines EHRs as: '*A comprehensive medical record or similar documentation of the past and present physical and mental state of health of an individual in electronic form and providing for ready availability of these data for medical treatment and other closely related purposes*.'

EHRs have the potential to be utilised for doctor–doctor communication as well as doctor–patient information (Blobel, 2004). Research has identified 4 possible models for the creation of and control over the EHR (Liaw *et al*, 1996):

- **Personal EHR model**: The patient is the chief manager and custodian of the record.
- **Shared EHR model**: The patient and doctor (GP) share the responsibility for maintenance and control of the record.
- **Trustee model**: The patient enters into a contract with a third party, the trustee, to keep and control the EHR.
- **Interoperable model**: This is a fully longitudinal system involving the transformation of the current paper records system into its electronic equivalent, operated at regional or national level with no patient involvement.

Countries around the world are adopting EHR models, which continue to expand across institutional boundaries with the development of EHR 'systems' (Rigby *et al*, 2007). In Ireland, the possible development of an EHR system has been the subject of much debate in recent times. The Department of Health and Children's (2004) *Health Information: A National Strategy*, for example, discussed the current domestic legal framework concerning health information and highlighted as concerns the lack of a system-wide framework for governance of health information and the lack of a coordinated policy for the archiving of personal health records or for the maintenance of organisational memory in health agencies.

[*] Article 29 of the EU Directive 95/46/EC establishes a 'Working Party on the Protection of Individuals with regard to the processing of Personal Data'. It is generally known as the 'Article 29 Working Party'. It is made up of a representative from the data protection authority of each EU Member State (including the Irish Data Protection Commissioner), the European Data Protection Supervisor and the EU Commission. The Working Party is independent and acts in an advisory capacity. The Working Party seeks to harmonise the application of data protection rules throughout the EU, and publishes opinions and recommendations on various data protection topics. It also advises the EU Commission on the adequacy of data protection standards in non-EU countries. For further information, *see* http://ec.europa.eu/justice/data-protection/index_en.htm

Although there are many potential advantages to having a functional EHR system in place, both for service users and service providers, it is also evident that many issues need to be considered at the outset, particularly those relating to consent, confidentiality and privacy.

One of the issues surrounding electronic health information is the concept of unique individual identifiers, or patient-identifiable data (CIHI, 2004). This issue is also the crux for the development of a data strategy. Patient-identifiable data refers to any personal data that can be used directly or indirectly to identify an individual. The concern as to why we need patient-identifiable data is addressed by Haynes *et al* (2007, p. 12), who argue '*updating, linking and validating data would be impossible without using some form of patient-identifiable data*'. Another issue is that research conclusions have the potential to be flawed if patient identifiers (such as age, gender, ethnicity, geographical location and socio-economic status) are not included.

However, in order to shift towards patient-identifiable data, many sources state that patients must give informed consent (Iversen *et al*, 2006). Haynes *et al* (2007) argue that obtaining informed consent from a large population may be prohibitively costly in terms of both time and money. They also make the point that patient-identifiable data are crucial in medical research and required for the developmental stages of disease registries. The need for fully informed consent, as prescribed in the Data Protection Act (1988), has the potential to bias the sample of who will be willing to be included in data collection and sharing. This issue will be addressed in more detail under 'The Irish legal framework: Data Protections Acts' below.

Staroselsky *et al* (2006) investigated the role patients could play in improving the accuracy of electronic health records. Patients were given access to their personal EHRs and allowed to amend information in their records. The outcome of the study found that patients can play a useful and necessary role in the accuracy of their personal data. For example, the documented number of people who had received the flu vaccine was quite low in the EHRs, but when patients were able to comment it appeared that the number was significantly higher, but had not been recorded on their records because they had received the vaccine from a source that does not feed into their records (Staroselsky *et al*, 2006). Although this study provides the potential for inclusion of patients in their personal EHRs, the authors continue to recommend the need for data sharing between and among healthcare providers.

Rigby *et al* (2007) highlight 4 challenges for the future of electronic patient record (EPR) systems:

- **The growth of expectation**: As patients continue to become more mobile and receive support from numerous agencies, there is a growing expectation that their EPRs will continue to be further integrated and 'follow them' throughout their lives. This means that EPRs will have to strive to further integrate data in the following ways: (1) by vertical integration, where all of the information for a patient is linked to one central location; (2) by horizontal integration, where information from various types of services will be fully integrated; and (3) by temporal integration, which will lead to a 'cradle to the grave' record. These dimensions of integration all play a role in the increasing expectation of the EPR.
- **Increasing digitised investigation and care delivery**: This essentially refers to the shift towards technology within the health sector and the need to be able to adequately capture each element of care and investigation, for example, the digitisation of X-rays, genetic information or wearable monitoring devices. Each of these has a role to play in the further development of the EPR.
- **Expanding person-focused 'health' boundaries**: This challenge reflects the change in understanding of health and its determinants. With the increasing awareness of the biopsychosocial model of health, it will be difficult for managers of EPRs to definitely decide what is to be included and what is not.
- **Dynamic environment of health**: This refers to the political structures that guide health and the various healthcare systems that exist worldwide. An understanding of the health environment is essential and will prove a challenge to the further development of the EPR.

Current Irish legal framework

Both the domestic and international legal framework need to be examined. To this end, sources such as the Constitution of Ireland, Acts of the Oireachtas, Statutory Instruments, EU Directives and EU law, the European Convention on Human Rights (ECHR) and finally the UN Convention on the Rights of the Child (CRC) are considered below.

(i) Constitution of Ireland

The Constitution of Ireland is inevitably the starting point in terms of the domestic legal framework. The right to privacy has been established as an unenumerated right under Article 40 in the seminal case of *McGee v Attorney General* (1974) and later in *Kennedy v Ireland* (1987). This right to privacy, however, has been established as 'not an unqualified right. Its exercise may be restricted by the constitutional rights of others, by the requirements of the common good and is subject to the requirements of public order and morality'.

(ii) Data Protection Acts

The main Irish law dealing with data protection is the Data Protection Act 1988. The 1988 Act was amended by the Data Protection (Amendment) Act 2003. The 2003 Amendment Act brought Irish law into line with the EU Data Protection Directive 95/46/EC on the 'Protection of Individuals with regard to the Processing of Personal Data and on the Free Movement of Such Data into Irish law'.

The Data Protection Acts are of central focus in terms of domestic Irish law since they set out the issues that must be considered concerning the processing of personal information relating to any individual. As the Data Protection Commissioner has advised, 'Any processing of personal data in an EHR system must recognise and incorporate the principles as set out in the Data Protection Acts'. The main principles that must be adhered to in these Acts may be summarised as follows:

- *Use limitation principle*: It prohibits further processing which is incompatible with the purpose(s) of the collection.
- *The data quality principle*: Personal data must be relevant and not excessive for the purposes for which they are collected and must also be accurate and kept up to date.
- *The retention principle*: Personal data cannot be kept for longer than is necessary for the purpose for which the data were collected or further processed.
- *Information requirements*: Data controllers processing information in EHR systems must provide certain information to data subjects, such as information on the identity of the controller, on the purposes of the processing, on the recipients of the data and on the existence of a right of access.
- *Data subject's right of access*: Data subjects can check the accuracy of the data and ensure that the data are kept up to date. These rights fully apply to the collection of personal data in EHR systems.
- *Security-related obligations*: Data controllers must implement appropriate technical and organisational measures to protect personal data against accidental or unlawful destruction or unauthorised disclosure. The measures can be organisational or technical.

Data relating to an individual's health is regarded as 'sensitive personal data' and so further protection is necessary to ensure that the data protection framework is conformed with. There is a general prohibition of the processing of personal data concerning health, as outlined in Article 8(1) of the Directive (*see above*), but there are a number of derogations to this:

- Article 8(2)(a): 'Explicit consent' – Consent must be 'freely given, specific and informed'.
- Article 8(2)(c): 'Vital interests of the data subject' – If the data subject is physically or legally incapable of giving his consent, the processing of data may take place in order to protect his or her vital interests.
- Article 8(3): 'Processing of (medical) data by health professionals' – this relates to providing health-related services of a preventive, diagnostic, therapeutic or after-care nature and for the purpose of the management of these healthcare services, e.g. invoicing, accounting or statistics.
- Article 8(4): 'Substantial public interest' exemptions – A Member State may introduce domestic legislation to enshrine further derogations to the prohibition of the processing of sensitive data provided that it is justified by reasons of 'substantial public interest'.

The right of access to personal data is subject to a limited exception in the case of health and medical records and in the case of social worker records, where allowing access would be likely to damage the physical, mental or emotional well-being of the individual. The Data Protection Commissioner also requires that any body or agency handling data relating to people's health or medical care needs to register with the Commissioner.

(iii) Freedom of Information Act 1997

The Freedom of Information Act 1997 allows an individual to access personal information held on record by public bodies, similar to the right afforded under the Data Protection Acts (*see above*). The requirements for public bodies as set out in the Freedom of Information Act may be summarised as follows:

- publish descriptions of their organisation, functions, records and services;
- publish policies, procedures and rules relating to access to services;
- provide access to records in accordance with the provisions of the Act;
- provide for correction of records containing personal information;
- provide findings of fact and reasons for decisions affecting an individual.

It is clear that there is a growing consensus that individuals have the right of access to this information and this factor must be considered in any Government proposals concerning personal data. The right of access by individuals to information about themselves, or 'personal access', is recognised as an essential element in the exercise of the right to privacy. This access empowers individuals to control their personal information and ensure its accuracy. As McDonagh (2006) comments, 'The recognition of such a right accords with the notion upon which modern data protection legislation is based, that personal information is the property of the individual to whom it relates'. The health professional or agency is the custodian of such information.

(iv) Statutory Instruments

Regulations were introduced in 1989 to restrict patient access to health information where this is likely to cause serious physical or mental harm. Such information can only be communicated following consultation with an appropriate 'health professional' (normally the patient's doctor). Similar Regulations were also introduced in relation to social work information where this is likely to cause physical or mental harm.

(v) Human Rights considerations

The Human Rights framework has seen the conceptualisation of the right to privacy and data protection as a human right. Such a development means that domestic Irish legislation must be re-examined in light of this. For the purpose of the present report, the European Convention on Human Rights (ECHR) and the UN Convention on the Rights of the Child (CRC) have been identified as the key instruments.

(a) European Convention on Human Rights

The incorporation of the European Convention on Human Rights (ECHR) into Irish law in 2003 means that the jurisprudence of the European Court of Human Rights must be considered. Two Articles in particular are of relevance for the purpose of this present discussion: Article 8, which sets out the right to privacy, and Article 10, which outlines the freedom of expression. As McDonagh (2006) points out, the scope of the right to freedom of expression as it applies to those who seek access to personal information tends to be interpreted more narrowly by the Court than the right to privacy. *Gaskin v United Kingdom* (1989) is an illustrative case in this regard as the applicant claimed that his rights under both Articles 8 and 10 had been violated. Gaskin sought access to information held by a public authority relating to a time when he had been in public foster care as he claimed that he had been abused during this period. The records in question included contributions from a number of professionals, some of whom had objected to their disclosure. In relation to Article 10, the Court held that there was not an obligation on the State concerned to impart the information in question to him, but held instead that the refusal to allow the applicant access to the records violated his right to respect for private life. Also noteworthy is that the Court held that the UK Government had breached the Convention by failing to provide for an independent authority with the power to decide whether access should be granted, in circumstances where a contributor to a record failed to give his or her consent to its disclosure, or withheld such consent.

This case clearly establishes the positive obligations on the State in relation to an individual's right of access to personal information and must be considered in any domestic Irish cases which raise similar issues.

(b) UN Convention on the Rights of the Child (CRC)

The UN Convention on the Rights of the Child (CRC) is examined in the context of the discussion concerning issues specifically relating to children and EHRs. Child law, on both the national and international level, clearly establishes that children's rights as regards consent, confidentiality and privacy must equally be considered. According to Schenk *et al* (2006, p. 94): '*The child rights' framework, in particular children's rights to participation and to protection, provides a powerful tool to guide decisions about designing data collection activities among children that can complement the principles of research ethics and broaden their application to programmatic activities … The use of participatory approaches emphasizes the importance of listening to children's views and creating opportunities for their meaningful involvement in data collection. On the other hand, the children's right to protection requires that they be protected from exposure to harm if, for example, collecting information about their personal circumstances may be considered upsetting or intrusive*.'

All of the above clearly points to a discussion of the CRC as a number of its Articles relate specifically to the legal issues under discussion in the present report. The great emphasis which the CRC places on the role of the family on the life of the child is evident throughout the text, most notably, at first glance, in the preamble. This emphasis on the role of the child's parents in particular must, however, be considered in the light of Article 5 of the Convention, which recognises the 'evolving capacities of the child'. This Article, coupled with Article 12, which recognises the child's right to be heard, are undoubtedly of importance in relation to processing information concerning the child.

Of utmost importance for this present discussion is Article 16, which outlines the child's right to privacy. It is clear from the literature that there is a constant tension between child protection and privacy issues, and again, it is important to reiterate that a balance needs to be struck between the two. Specifically, a child's right to privacy, according to Anderson *et al* (2006), is '*the mechanism by which we define who we are in relation to other people, and as such can be seen as an essential element of child welfare and child protection because it encourages the development of clear personal boundaries … It is thus important that adults maintain a scrupulous respect for privacy in their dealings with children, in order to reinforce personal boundaries and underline each child's right to say "no" to unwanted intrusion. In this way, the right to privacy directly empowers children to protect themselves*'.

Legal issues regarding development of electronic health records in the UK

This section focuses on the development of electronic health records (EHRs) in the UK and identifies the relevant literature, discusses some of the key issues and asks what lessons can be learned from the experience there.

In the UK, as elsewhere, all patients have the right to privacy and the reasonable expectation that the confidentiality of their personal information will be maintained by all healthcare professionals and this is so regardless of the form of the information, be it paper, electronic, photographic or biological (European Standards on Confidentiality, 2005, p. 7). Introducing an electronic health records (EHR) system clearly raises many legal questions. As the report by the Article 29 Data Protection Working Party (2007, p. 5) points out, '*EHR systems introduce a new risk scenario, changing the whole scale of possible misuse of medical information about individuals*'. The main information governance issues and developments that need to be addressed in relation to such records are summarised by the UK Department of Health as follows (with detailed discussion of each below):

- consent;
- anonymisation and pseudonymisation;
- data ownership, control and access;
- research;

- electronic communication and information governance;
- role-based access control and smartcards;
- other systems issues.

The Article 29 Working Party's series of 'special safeguards' that would be necessary to protect the individual's data protection rights when considering the establishment of an EHR system are also worth noting at the outset and can be summarised as follows:

- respecting self-determination;
- identification and authentication of patients and healthcare professionals;
- authorisation for accessing EHR in order to read and write in EHR;
- use of EHR for other purposes;
- organisational structure of an EHR system;
- categories of data stored in EHR and modes of their presentation;
- international transfer of medical records;
- data security;
- transparency;
- liability issues;
- control mechanisms for processing data.

Consent

It is clear that informed consent is necessary as regards the processing of confidential patient information, unless such processing is otherwise set out in law or necessary for the public good. Consent can be expressed or implied, but the patient must be informed. The patient's self-determination concerning when and how his or her data are used is vital according to the Article 29 Data Protection Working Party, which suggests that agreement as a safeguard may be given in the form of an opt-in or an opt-out and there should be different degrees of the possibility to exercise self-determination. Furthermore, it should always be possible for a patient to prevent disclosure of his medical data and the question of possible complete withdrawal from an EHR system needs to be addressed. The Article 29 Working Party also submits that transparency is necessary in relation to the content and the functioning of an EHR system in order to ensure trust and confidence in the system. Notification to Data Protection supervisory authorities combined with special information, which is easily available and understandable, must be procured by the controller of the system. The Article 29 Working Party also suggests that the Internet may be the ideal information distributor in relation to EHRs and that free-of-charge, easy-to-use, but safe access points for data subjects to check on the content and on disclosure of their EHR record might also be useful in terms of gaining trust in the system.

The Information Commissioner's Office (ICO) is the UK's independent authority set up to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals. Its website (www.ico.gov.uk/) offers a useful guide on issues relating to the child and consent, an area which is by no means straightforward and instead needs to be treated on a case-by-case basis. Section 2A(1) of the Data Protection Act 1998 addresses consent and sets out that 'consent cannot be obtained from a person who, by reason of age, is likely to be unable to appreciate the nature and effect of such consent'. Despite this statement, the legislation does not put forward a minimum age at which consent of a child can be obtained.

In the Medical Law field, a young person of 16 years can consent to any issues concerning their health. In the case of a person under 16 years, the relevant medical practitioner must draw on his or her respective judgement in assessing whether the child in question has the maturity 'to understand and make their own decisions about the handling of their personal health information'. In relation to the right of access to health data for children under 16 years, the relevant medical professional must again decide and he or she may allow such access to (i) the individual alone; (ii) a parent or guardian alone; or (iii) both jointly. The maturity of the child, as well as his or her best interests, are again key concepts that need to be considered.

Anonymisation and pseudonymisation

Anonymisation may be described as 'the removal of name, address, full post code and any other detail or combination of details that might support identification'. Pseudonymisation of data differs from anonymised data as the original provider of the information may retain a means of identifying individuals. Data that cannot identify an individual patient, either directly or through linkage with other data available to a user, do not need to be regarded as confidential. The Department of Health in the UK sets out that:

> 'As a general rule, for purposes other than direct care or the quality assurance of that care, it is advisable to work to the principle that:
> (a) wherever possible anonymised information will be employed,
> (b) that the use of pseudonymised information will only be considered where anonymised information cannot satisfy requirements, and that
> (c) patient-identifiable information will only be made available where neither of the other categories can provide what is needed and it is lawful to do so.'

The website of the Information Commissioner's Office, UK (www.ico.gov.uk/) details that a medical professional can pass patient data to various health bodies for administrative purposes if that data are anonymised or aggregate data, from which individual patients cannot be identified. Patients should, however, ideally be informed in advance of such uses of their personal data. The patient's consent is required in situations where the personal data concerned includes identifying details. As regards the international transfer of medical records, the Article 29 Working Party states that such data can only be transferred to countries outside the EU in anonymised form or at least in pseudonymised form. Any processing of such data should take place within jurisdictions applying the EU Data Protection Directive or an adequate data protection legal framework. In relation to clinical studies, secure pseudonymisation must be required as a minimum prerequisite.

Data ownership, control and access

The Data Protection Act 1998, which incorporates the EU Data Protection Directive 95/46/EC, is the main governing legislation in the UK. The patient is the 'data subject' and the health professional is the 'data controller'. The National Programme for Information Technology has set about changing the current structure whereby data are held locally so that data will instead be controlled by the Department of Health and the NHS. In relation to the right of access to health data for children under 16 years, the relevant medical professional must decide and he or she may allow such access to (i) the individual alone; (ii) a parent or guardian alone; or (iii) both jointly. The maturity of the child as well as his or her best interests, are again key concepts that need to be considered.

Article 6(1)(c) of the Data Protection Act 1998 sets out that only relevant information should be entered into an EHR. Different data modules with different access requirements could satisfy any confidentiality issues. In order to satisfy any private insurance companies' entitlements to certain health records, it may be necessary to establish a standardised special 'documentation package'.

Research

The UK Department of Health operates under the general principle that no disclosure of data should be allowed without the approval of the relevant patients, clinicians and research ethical committee(s). There may, however, be legitimate reasons for extracting patient-identifiable data from a GP system, other than for routine clinical care, but this will need to be subject to the informed consent of the patient or the Secretary of State and also be with the knowledge and informed consent of the guardian of the record (in this case, the GP) and follow approval from a Research Ethics Committee. The Department of Health furthermore suggests that there should be both an audit trail for the data extraction and retention of the research database in order for both patients and health professionals to satisfy themselves, if necessary, that the data have been handled ethically and legally.

The Data Protection legislation allows the use of personal data for research or statistical purposes in situations where the patient in question was not informed in advance, provided that no damage or distress is likely to be caused to the individual. The Health (Provision of Information) Act 1997 provides, however, that any person may provide any personal information

to the National Cancer Registry Board for the purpose of any of its functions, or equally to the Minister for Health or any body or agency for the purpose of compiling a list of people who may be invited to participate in a cancer screening programme authorised by the Minister.

Electronic information governance

Clinical messaging: The scope of clinical messaging is to be extended to include facilities to request and receive reports for the full range of laboratory and diagnostic imaging procedure; to receive notifications of hospital admission, of casualty and of OOH attendance; electronic transfer of prescriptions from GP practices to pharmacies; and GP to GP electronic transfer of records.

NHS e-mail 'Contact': The current version of NHS e-mail ('Contact') provides security for messages sent between two Contact e-mail addresses. Patient-identifiable information can be safely sent from one Contact e-mail address to another. If either the sending or receiving address is not a Contact address, then separate encryption will be needed for sending confidential information, including Patient-Identifiable Data.

Role-based access control and Smart cards

The Article 29 Working Party proposes that the identification and authentication of patients and healthcare professionals through health cards or Smart cards could be used. Smart cards provide enhanced security and controls over access to the systems and the messages they send. They carry the user's roles and therefore levels of access to systems. Access to electronic patient records will depend on both the establishment of a legitimate relationship between the clinician and the patient, and the use of a Smart card and PIN. The Article 29 Working Party further suggest that authentication by means of electronic signature could be used in conjunction with the cards. Finally, only those healthcare professionals who presently are involved in the patient's treatment may have access. Modular access rights are also suggested, meaning that categories of medical data in an EHR system could be formed and these limited to corresponding categories of healthcare professionals or institutions. It is also mooted that patients should be given the chance to prevent access to their EHR data if they so choose.

Liability issues

Possible infringements of privacy must be balanced by liability for damages caused. The Article 29 Working Party suggests that any Member State wishing to introduce an EHR system should in advance carefully conduct in-depth, expert civil and medical law studies and impact assessments to clarify the new liability issues likely to arise in this context regarding, for example, the accuracy and completeness of data entered in an EHR, defining how extensively a healthcare professional treating a patient must study an EHR, or about the consequences under liability law if access is not available for technical reasons. The report by Korin and Quattrone (2007) entitled *Electronic Health Records raise new risks of malpractice liability* outlines some of the risks of malpractice liability associated with EHRs, which include data loss or destruction, inappropriate corrections to the medical record, inaccurate data entry, unauthorised access and errors related to problems that arise during the transition to EHRs.

This report suggests that the EHR system should provide a clear record of when a correction or addendum is made in accordance with facility policy. As the report details, the integrity, accessibility, security and privacy of protected health information must be addressed through policies and procedures that support the requirements of security and privacy. EHR systems may also have built-in safeguards to flag certain kinds of data entry errors and to prevent inadvertent data loss.

This report further raises the question as to whether medical practitioners will have a legal duty to access patients' past medical records, which historically has been adjudicated on a case-by-case basis, and concludes that a standard of accepted practice will evolve in this respect. A further issue to be considered is that medical practitioners may choose a course of action that is not contemplated in a clinical guideline provided in the EHR system and so a computerised prescriber order-entry system that requires users to document reasons for clinical overrides may constitute documentary evidence which could be used in a malpractice case, for example. The report goes on to list a number of other possible legal questions that could arise concerning this

scenario relating to the standard of care and the role of software vendors and manufacturers in such cases:

> 'Will the standard of care be shaped by a software vendor's choice of clinical-decision-making tools? Might software vendors or software manufacturers routinely become co-defendants and/or witnesses in medical negligence lawsuits? Will greater weight be assigned to electronically generated guidelines that are built into system software created by a healthcare institution rather than provided by a commercial vendor? Can a physician successfully assert a defence that "alert fatigue" caused him/her to override an alert because the system routinely provides overrides that clinicians view as inappropriate?'

Further issues, such as the failure of a medical professional to diagnose and treat patients in a timely manner, may arise due to the level of information that the EHR may contain. EHRs may also raise the cost of litigation because of the need for expert testimony in the fields of health informatics or health IT.

Control mechanisms for processing data in EHR systems

The Article 29 Working Group lists some possible mechanisms that could be used to evaluate the existing safeguards:

- a special arbitration procedure for disputes concerning the use of data in EHR systems;
- a single special institution must be made responsible towards the data subjects for the proper handling of access requests;
- a special routine for informing the data subject when and who accessed data in his EHR could be introduced;
- regular internal and external data protection auditing of access protocols must take place.

Other systems issues

- **Risk management**.
- **Accessibility**: On the practical level, practices need to ensure that they have an adequate number of workstations at each point within the organisation where staff can access the EPR or other supporting applications.
- **Capacity and storage**: The system must have adequate data storage capacity to meet storage needs.
- **Physical security**: The system must be safe and secure – security measures to prevent loss or failure due to theft, fire, power failure and computer viruses, for example, need to be taken. In this regard, back-ups need to be performed regularly and stored securely.
- **Access control**: The use of Smart cards and role-based access control are necessary and until their implementation, practices must be undertaken to ensure controlled access.
- **Security policy**: The Article 29 Working Party sets out that access by unauthorised persons must be virtually impossible and prevented. Privacy-enhancing technologies (PETs) should be applied. Encryption should be used for both the transfer and storage of data in EHR systems. All security measures should be constructed in a user-friendly way to broaden their application. The Working Party's report sets out a number of security measures that must be adhered to, in addition to all of the above:
    - the development of a reliable and effective system of electronic identification and authentication, as well as constantly updated registers for checking on the accurate authorisation of persons having or requesting access to the EHR system;
    - comprehensive logging and documentation of all processing steps which have taken place within the system, especially access requests for reading or for writing, combined with regular internal checks and follow-up on correct authorisation;
    - effective back-up and recovery mechanisms in order to secure the content of the system;
    - preventing unauthorised access to or alteration of EHR data at the time of transfer or of back-up storage, e.g. by using cryptographic algorithms;
    - clear and documented instructions to all authorised personnel on how to properly use EHR systems and how to avoid security risks and breaches;
    - a clear distinction of functions and competences concerning the categories of persons in charge of the system or at least involved in the system with a view to liability for shortcomings;
    - regular internal and external data protection auditing.

- **Disposal**: All storage media should be re-formatted to delete any personal information before disposal. If there is any possibility that such information might remain accessible on the storage medium after formatting, then the hardware should be physically destroyed before disposal.
- **Disaster recovery**: A detailed disaster recovery plan is necessary for health professionals before moving to paperless practice and should include the following:
    - Back-up of the system to a suitable medium (usually magnetic tape) at regular intervals, with a frequency of no less than once per day;
    - a system of cycling multiple media such that a single failed back-up cannot render the plan ineffective (e.g. using different tapes for each day in a weekly cycle);
    - secure storage of back-up media to protect against accidental damage (e.g. flood or fire) or theft;
    - a system to ensure that at least one recent back-up is retained off-site to provide additional resilience against accidental destruction or theft (e.g. taking the previous day's back-up off-site each evening);
    - a system to ensure that any warnings or messages produced by the back-up system are noted and acted upon;
    - regular replacement of back-up media in accordance with the manufacturer's instructions;
    - periodic submission of a specimen back-up to an external verification service (where available) to ensure that back-ups obtained are able to be used to restore a functioning system.

Data protection

The European Data Protection Act 1998 sets out requirements for the use and handling of personal data for all Member States of the European Union. The major requirement of the Act is that informed consent must be rendered prior to personal data being shared between organisations. However, the Act makes special provisions for the research arena. Lawlor and Stone (2001, p. 1221) indicate that '*data may be exchanged between organisations for the purpose of research providing that the research is not used to support measures relating to particular individuals or could cause damage or distress to individuals*'. With respect to implications for public health practice and research, it appears that the exchange of fully anonymised data for the purpose of public health practice is permitted under current European legislation. The majority of the public debate surrounding data protection appears to be concerned with confidentiality. Lawlor and Stone (2001) argue that there is little attention placed on the detrimental effects that data protection legislation may have on data sharing.

International examples of data protection specific to children are given below.

France

Data protection in France is governed by the Law on Informatics, Files and Freedoms of 1978, as amended in 2004 to incorporate the EU Data Protection Directive 95/46/EC. The National Commission for Informatics and Freedoms (CNIL) is the data protection authority. French law requires that processing of personal data in the public sector must either be specifically regulated by a decree adopted after the opinion of the CNIL was obtained, or authorised by the CNIL itself. The regulatory Acts specify in detail the specific purpose of the processing covered by the regulatory Act, the categories of data concerned and the categories of recipients who may have access to the data, any restrictions on rights of data subjects and so on. Controllers can also choose to process data in accordance with general 'simplified norms' issued by the CNIL for a particular sector or field of activity.

Children and information sharing

In a 2001 report on *The Internet and the collection of personal data from minors*, the CNIL discussed some issues central to the question of data protection and minors generally, including the nature of a child, the different ages at which young people are competent to take certain decisions in different contexts, the importance of families and parents, and the Commission's

general approach to the collecting of data from minors. The CNIL stressed in that report that '*the guarantees provided to anyone by [the French data protection law] must be applied with special force as concerns minors*'. This means that while there may be special justification for the collecting of data on children in the public sector, this should be subject to strict purpose-specification and limitation (including clear and narrow purpose-definitions); parents should be fully informed and in principle asked for their (written) consent for the processing of data on their children; data on siblings and parents should not be obtained from children; and the collecting and further use (and, of course, especially the disclosure) of sensitive data on children should be particularly strictly circumscribed, in rules drawn up by, or drafted following the advice of, the CNIL.

In 1981 in France, the GAMIN system was proposed, which was intended to involve the automated processing (and analysis) of data from health certificates, which by law have to be drawn up by a doctor on any new-born child and its mother. The CNIL, however, prevented the system from becoming established on the grounds that it violated the basic data protection principles. The system was to identify children, via computer-generated 'alerts', who are in need of special attention by social services, for example, on the basis of heterogeneous data. In France, the law requires that 'human identity' should never be reduced to a computer model and no significant decisions should ever be reached in this way. This applies regardless of whether it involves computer analysis of data held by just one agency or by several. The CNIL subsequently approved a more limited, experimental system for ten departments over three years, which allowed the creation of two separate files (without links) and the disclosure of limited information from medical certificates to social services in certain circumstances – but, notably, without the use of any computer-generated 'flags of concern'. In 1985, the CNIL adopted a general recommendation on the collection of personal information of children in schools, requiring that questionnaires not be given to school children without the prior written consent of the parents and that the parents' written consent relating to the dissemination of photographs of minors on the Internet also be sought.

The rules on the system for collecting and storing of personal data on students at secondary schools and academies (the 'SCOLARITÉ' system) strictly limit the amount and nature of the data that can be collected by these institutions, the recipients of the data and the uses that can be made of them. They are particularly strict in the limitations imposed on the uses that can be made of the data by the central (national) authorities. It is clear from these rules (and from the GAMIN-ruling) that the CNIL would strongly oppose the adoption of any specific provisions allowing disclosure of data on students in order to generally identify (through computer analysis or otherwise) whether any of them are likely to need special attention by social services or other child protection bodies. The CNIL would, in any case, be extensively consulted on the drawing up of any such specific rules. In France, general catch-all provisions would not suffice to replace specific provisions.

In conclusion, it is clear that the data protection authority in France, the CNIL, plays a central role in both formulating and subsequently monitoring the data protection rules in France. The special recognition which it affords to children and their respective data protection and privacy rights means that data collected and shared must be subject to the following:

- strict purpose-specification and limitation;
- parents must be fully informed and in principle asked for their consent for the processing of data on their children;
- data on siblings and parents should not be obtained from children;
- the collecting and further use and disclosure of sensitive data on children should be strictly circumscribed in rules drawn up by, or drafted following the advice of, the CNIL.

Germany

In Germany, legislative power is divided between the Federation and the States. The legal framework concerning data protection in Germany, therefore, needs to be examined at local, regional and State level, as well as at the Federal level. There are 16 States in Germany, resulting in some 17 general data protection laws, including the Federal Law and one for each of the States. There are also other laws that deal with data protection in a specific context or contain certain rules that need to be considered. Each State has a data protection commissioner and there is also a commissioner at Federal level.

Despite all of this, the laws in each State embody the same basic principles as set out in the German Constitution. In the *Census* judgment, the German Constitutional Court (1983) established a 'right to informational self-determination' from the 'right to respect for one's personality' as set out in §2(1) of the Constitution. Subsequent cases have seen this develop into a right for individuals to know (or to find out) who collects data on them, when and for what purposes; and that strict limits are placed on the collecting, storing, use and disclosure of personal data. Particular emphasis is given to the principle of purpose-specification and limitation; the purpose in question needs to be defined narrowly. The principle of data minimisation means that only the minimum amount of data necessary for the purpose may be collected and held. Personal data must, in principle, be obtained from the data subject him or herself, rather than from other sources. Finally, because collecting, storing, using and/or disclosing data all constitute interferences with a constitutional right, there must be a statutory basis for such actions.

Specifically in relation to children, the German Constitutional Court has ruled that children need special protection and the collecting, storing and usage (and disclosing) of data on minors deserves greater justification and general data protection principles should be strictly applied.

Constitutional data protection principles for children

Consent can only be given for defined purposes and provided that the person concerned could properly judge the implications of giving his or her consent. In principle, consent must be given in writing and in the case of sensitive data there must be specific authorisation. The question of free consent is particularly important and German data protection law requires that all circumstances be taken into account, including the relationship between the data subject and the body seeking the data; the nature of the data; the uses and disclosures for which consent is sought and their proximity to the relationship between the data subject and the body seeking the data; the importance and possible effects of the processing for which consent is sought for the data subject; and finally, the capacity of the data subject to appreciate the importance of his or her consent. In Germany, a person is considered an adult at 18 and is then competent to decide on matters concerning consent. The age of 14 is, however, viewed as an important threshold in relation to minors' capacity to consent, although 16 years is the more usual age in this respect.

Constitutional principles relating to the disclosure of data

In the public sector in Germany, data may only be disclosed by one public body to another either with the express, valid consent of the data subject or if a law either requires or allows such a disclosure. As outlined in the report *Children's Databases – Safety and Privacy* by Anderson *et al* (2006), conducted for the Foundation for Information Policy Research, limited exceptions exist with regard to the establishment of 'joint processing operations' and 'on-line processing operations', as well as for centralised 'shared [Federal/State] databases'. Shared Federal/State databases are extremely rare and in practice, until now, limited to Federal/State police purposes only.

Applying the constitutional principles in practice: Schleswig-Holstein

As Anderson *et al* (2006) outline, the various agencies working with children – such as schools and education services, child welfare, social and health workers, and the police – in principle work separately on each child. Each agency needs data on the respective child in order to perform its respective task and therefore needs a special statutory basis for its personal data collection, use and disclosure. These agencies must equally comply with the general State and Federal data protection law.

In relation to disclosures of data from one public body to another, §14(1) of the State data protection law (LDSG) allows this to take place:

(i) when the data subject has given his or her consent;
(ii) when the LDSG or another law expressly allows it;
(iii) when the data are 'necessary' for the fulfilment of a statutory task of the disclosing or the recipient body; or
(iv) to protect 'vital interests' of the data subject (e.g. when the data subject is in coma in hospital).

The rules on data sharing arrangements are set out in §8 of the State data protection law. They apply to:

- automated processing operations, which allow several bodies to jointly process personal data ('joint processing operations');
- automated processing operations, which allow several bodies to disclose data online ('on-line processing operations').

Such operations may, however, 'only be established insofar as their establishment is proportionate, taking into account the legitimate ['protection-worthy'] interests of the persons concerned and the tasks of the official bodies concerned'. As Anderson *et al* (2006) point out, this prior check can, in principle, be carried out by an in-house data protection official, appointed by a public body to ensure compliance for its operations with the relevant data protection laws and regulations. In Schleswig-Holstein, the Data Protection Commissioner can effectively prohibit the establishment of a joint or online arrangement if he or she feels that it poses a disproportionate risk to the rights and interests of the data subjects.

It is clear that data sharing arrangements in Germany have to comply with a number of important requirements. For joint operations, the 'processing particulars' that have to be drawn up (and usually notified to the Commissioner) must include also 'the areas of processing for which each of the parties to the arrangement is responsible'. This means that it must be clear which body is responsible for which part of the processing operation. However, data subjects can contact any of the bodies involved in the exercise of their rights (such as the right of access, correction or erasure). If joint processing operations lead to the disclosure of data (from one of the parties involved to another, or from one of them to a body that is not party to the joint arrangement), the recipient, the time of disclosure and the data that are disclosed must be recorded and the record of such disclosures must be kept for one year. With regard to online operations (i.e. online accessible databases), anybody that accesses (and downloads) data is responsible for ensuring that this action complies with data protection rules (in particular, that the data are necessary for the performance of the task of the body in question). The body that uploaded the data is not responsible, except that it should carry out checks on the permissibility of the downloading if it finds reasons to do so. However, it is required to carry out spot-checks to see if there is such cause. As Anderson *et al* (2006) note, however, 'in practice no such joint or online systems have been established, or are likely to be allowed to be established, in the areas of interest to this study. Data disclosures and exchanges in these areas must therefore be assessed under the general rules on disclosures … and (more importantly) under the special data protection rules in the special laws covering the specific activities concerned'.

The report by Anderson *et al* (2006) focuses on data on school children as an example and notes that the rules governing this are set out in 'data protection in education' in the Schleswig-Holstein School Law and also in a School Data Protection Regulation, which includes an annex listing the data that schools may collect and process in further detail. The only data that may be held by schools are on the parents of children, including name, address and telephone number. The results of 'school, medical, school-psychological, or school-pedagogic examinations' (i.e. examinations by the doctor, psychologist or educational experts attached to the educational service) may be kept on file, but may 'under no circumstances' be entered on a computer.

Interagency cooperation and the disclosure and/or sharing of data

The disclosure and sharing of data between various agencies is subject to two main considerations: (1) there must be a capacity to disclose data by one agency and to collect and store data by the other; and (2) that this is done on a case-by-case basis, thereby assessing the needs for such arrangements. The basic data protection principles should be taken into account in this respect.

The Data Protection Commissioner in Schleswig-Holstein prepared a paper outlining the data protection requirements relating to the cooperation between schools and the official youth care agencies (*see* www.datenschutzzentrum.de/material/recht/ldsg-eng.htm). The consent of the student in question is vital since 'Only by involving the person concerned can trust be created or maintained, in relation to the school, to the youth care team or to other involved parties, which is required to provide effective help [to that person]'. In relation to the age of the student and his or her capacity to consent, the Commissioner's report continues:

> ‘The assessment of the competence [of the child or young person] to assess this matter can only ever be made on a case-by-case basis, taking into account all the circumstances (age of the young person, psychological maturity, scope of the data processing, as well as scope, aim, recipient, time-period, sensitivity of the procedures). If a person under 18 years old can be deemed capable of making the decision, this overrides contrary wishes of parents or carers. However, if the latter have expressed such contrary views, a particularly careful assessment must be made, since this implies that the [proposed] co-operation between the school and the youth care agency has an effect on the relationship between the youth and his or her guardians. As far as children under the age of 14 are concerned, it has to be assumed, as a rule, that they do not yet have adequate capability to assess the often complex issues.’

The Commissioner further emphasizes that:

> ‘Cooperation between a school and youth care agency in relation to a specific [young] person can only ever take place in a specific case. The necessity of data exchanges [implicit in such co-operation] must be assessed with reference to each child or young person. General collecting of personal data from schools for tasks of the youth care agency, e.g. to determine the need for extra-curricular help, is not permitted.’

The Commissioner gives examples of when a school may give information to a youth care agency, at the request of that agency, for example, if the child is in an emergency situation and the agency needs the information to take the child into care or if the agency has been asked by a Court to provide a report on a child that requires such information. Equally, the Commissioner lists some possible scenarios that would enable a school to approach a youth care agency and pass on information on a child, for example, if the school had well-founded reasons to suspect that the child is abused or neglected, or if the child is in an emergency situation in which the agency can help.

The Commissioner also prepared a report on the data protection requirements for cooperation between youth care agencies and the police. Again, the need to balance interests in each individual case was emphasized:

> ‘Passing on of data [on a child or young person] can deter the person concerned from seeking further help from the youth care agency. It is quite possible that the agency was contacted precisely in order to resolve or lessen the problems with an (offending) young person without involving the police. Often, and in particular in relevant parts of society, there is a fundamental reluctance to engage with the police. The assurance of confidentiality often plays a big part in the success of the help offered. Passing on information to the police will often undermine the chances of success.’

The main principles concerning data sharing in Germany in relation to children may be summarised as follows:

- professionals involved in youth care are under a very heavy duty of confidentiality in respect of any personal data they collect on their young clients, and especially in respect of any data that are ‘entrusted’ to them;
- that obtaining data by any public body, disclosure of data by any public body to another public body, and receiving and further use of the data by the other public body requires a specific legal basis; and that the confidentiality rules relating to ‘social data’ entrusted to youth care workers and teachers override more general rules allowing other bodies – including the police – to request or even demand information;
- that this does not prevent disclosures of personal data on children and young persons if this is necessary to protect them from harm, or if the disclosure does not undermine the providing of assistance to the children and young persons concerned;
- that the decision on whether to disclose data on a child or young person is ultimately always left to the youth care professional: he or she should decide to disclose or not to disclose data on the sole basis of what is in the best interest of the child.

Other European perspectives

The 2006 report *Children's Databases – Safety and Privacy* by Anderson *et al* also discusses some further issues pertaining to data sharing in various jurisdictions and of particular relevance for the purpose of this discussion is that relating to the use of general identifiers in data sharing. Article 8(7) of EU Data Protection Directive 95/46/EC sets out that: '*Member States shall determine the conditions under which a national identification number or any other identifier of general application may be processed*'. Anderson *et al* (2006) provide some examples:

- In Finland, the use of the national identity number is generally allowed with the consent of the data subject, but imposes strict limitations on its use without consent.
- In Denmark and the Netherlands, exchanges of this national identification number between public bodies, also without consent, occurs where this is considered useful for the work of the bodies in question.
- In Sweden and France, the use of the national number, even with the consent of the data subject, must still be 'clearly justified'. In France, for example, the CNIL has sought to limit the use of the number to clearly specified circumstances and has attempted to prevent the use of the number for the creation of interconnections between databases operated by different (mainly public sector) bodies.

On this note, Anderson *et al* (2006) further point out that in many EU countries, the linking of databases, and/or the use of general identifiers that facilitate such linking, are regarded as posing inherent risks to the rights and freedoms of data subjects that should be addressed by special, strict substantive and procedural rules. In Sweden, Denmark and the Netherlands, for example, the general identifier can only be used to facilitate links between different data collections if the disclosures and sharing involved is justified ('clear justification'). In Austria, Greece, Luxembourg, Finland, Portugal, France and Germany, all 'interconnections' between databases and all data matching which arises as a result require a permit from the Data Protection Authority, or are subject to the requirement of a 'prior check' or 'prior authorisation'.

Ethical considerations

The European Commission funded the EuroSOCAP Project to consider the privacy and confidentiality of healthcare information and as a result, produced a set of standards and guidance to inform healthcare professionals and healthcare provider institutions as regards best practice in protecting patients, particularly vulnerable patients, throughout the healthcare sector of the European community. While these standards also incorporate the relevant European legal provisions as set out in the European Convention on Human Rights and the EU Data Protection Directive 95/46/EC, it is important to note, as McClelland and Harper (2007) point out, that:

> 'Such laws do not exhaust the obligations on healthcare professionals to respect and protect patient confidentiality and privacy. Healthcare professionals may also need to exercise professional judgment. These Standards provide ethical guidance to all healthcare professionals in the making of such judgments. Best ethical practice also requires a supportive context and the Standards contain recommendations to healthcare provider institutions on those measures necessary for the most effective realisation of the Standards in practice.'

Thus it is clear that ethical standards may differ from legal standards in a particular jurisdiction. In this regard, it is worth noting that health professionals are obliged to follow their ethical obligations where such standards require greater protection for patient confidentiality and privacy than the legal standards. Vulnerable patients, such as children, have greater needs as regards confidentiality and healthcare professionals have a duty to recognise this. As Article 8 of the Universal Declaration on Bioethics and Human Rights (2005, *see* www.unesco.org/new/en/social-and-human-sciences/themes/bioethics/bioethics-and-human-rights) sets out:

> 'In applying and advancing scientific knowledge, medical practice and associated technologies, human vulnerability should be taken into account. Individuals and groups of special vulnerability should be protected and the personal integrity of such individuals respected.'

An example of a patient's vulnerability in relation to information sharing is the lack of decision-making capacity as set out in law. To counteract this, the EuroSOCAP Project recommended that

a clear institutional framework be established to protect those who lack decision-making capacity in relation to the protection, use and disclosure of their confidential patient information.

Medical confidentiality is a core value of European healthcare and it consists of the following:

- Individuals have a fundamental right to the privacy and confidentiality of their health information.
- Individuals have a right to control access to and disclosure of their own health information by giving, withholding or withdrawing consent.
- For any disclosure of confidential information, healthcare professionals should have regard to its necessity, proportionality and attendant risks.

The professional's duty of confidentiality is therefore not absolute, given the wider contexts that must also be considered.

Key ethical concerns

The following key ethical concerns need to be addressed in relation to electronic health records (EHRs):

- Healthcare professionals must respect patients' requests for access to their healthcare information and comply with their legal obligations under data protection laws.
- Healthcare service providers must ensure that there is an active, effective and appropriate policy about informing patients and/or their legal representatives in each setting about the protections, uses and disclosures of their information.
- Healthcare professionals must ensure that patients and/or their legal representatives are informed in an appropriate manner:
    - of what kind of information is being recorded and retained;
    - of the purposes for which the information is being recorded and retained;
    - of what protections are in place to ensure non-disclosure of their information;
    - of what kinds of information sharing will occur;
    - of the choices available to them about how their information may be used and disclosed;
    - about their rights to access and where necessary to correct the information held about them within healthcare records;
    - of the information required to be provided to them by national law implementing EU Data Protection Directive 95/46/EC;
    - of country-specific legal provisions or principles governing disclosure.
- Patients, or where appropriate their legal representatives, must be informed of what information sharing is necessary for the patient's individual healthcare. Provided they are informed in this way, explicit consent is not necessary; implied consent is sufficient for the ethical sharing of patient information for their healthcare.
- Provider institutions must ensure that the express consent of the patient (or of their legal representative) is obtained for processes of clinical audit by staff not involved in the care of that patient. Where it is proposed to make information available outside the health provider institution, the audit process should also be subject to ethical review.
- Healthcare professionals should strive to ensure that institutional policies for clinical audit are compatible with the ethical requirement for confidentiality.
- All organisations providing healthcare should ensure that all people employed by or working in the organisation are under a legal obligation to protect patient confidentiality.
- The potential benefits of information sharing with their carer should be discussed with the patient and/or their legal representative. However, the fact that such information sharing may be beneficial does not diminish the duty of confidentiality owed to the patient by the healthcare professional.
- Service providers must establish and ensure the adoption of clear publicly accessible protocols for information sharing within teams, beyond teams and with outside organisations.
- Healthcare professionals may have different criteria and thresholds for the disclosure of confidential information, for example, in relation to public safety. It is essential that each healthcare professional familiarise him or herself with such differences and moderate disclosures accordingly.

Challenges to the development and use of electronic health records

In summary, there are numerous challenges that have the potential to compromise the effectiveness of electronic health records (EHRs). The first is the growth of expectation. This means that users and organisations will expect that EHRs become as mobile as the individuals to whom they belong. There is currently a need to integrate information between organisations and agencies. Rigby *et al* (2007) highlight 3 dimensions that have the potential to further integration:

- *Vertical integration*: Integration of primary care and secondary care records into one single shared view of the individual.
- *Horizontal integration*: A shift from organisational records to regional records.
- *Temporal integration*: Further push towards electronic records in order to develop the potential for a 'cradle to the grave' patient record system.

The second challenge facing EHRs are expanding person-focused 'health' boundaries. This means that there is a greater recognition of the factors influencing health that exist outside of the healthcare setting. This calls on different types of data beyond administrative data that are typically generated within the primary care setting. With the expansion of 'health' boundaries, there is more emphasis on integrating the variety of information. Finally, the dynamic environments in which individuals live provide challenges to the concept of EHRs.

4. Examples of data strategies

The Netherlands

Statistics Netherlands is an autonomous agency and acts as the central bureau of statistics in the Netherlands, responsible for collecting and processing data in order to publish statistics to be used in practice by policy-makers and for scientific research, as well as producing European (Community) statistics (*see* www.cbs.nl/en-GB/menu/home/default.htm). Statistics Netherlands today operates as a '*single standardised production line for all statistics with a central data repository*'.

Heerschap and Willenborg (2006) provide a detailed overview of the shift in statistical systems that occurred in the Netherlands during the 1990s. Statistics Netherlands had recognised that its current statistical system was not proficient enough to meet the demands being placed on it. In particular, it was the Division of Business Statistics that initiated the change, with the main goal of developing one overall architecture for all statistics produced within the department. This was to include merging social statistics, business statistics and national accounts. International discussions on statistical procedures shaped the process undertaken. For example, strong emphasis on centralised metadata proved to be a recommended strategy.

In order to fully appreciate the current approach in the Netherlands, it is important to take note of the process involved in reaching it. The former situation in the country was described by Heerschap and Willenborg (2006) as a 'pure product stovepipe model'. In essence, this meant that each piece of research commissioned was viewed as a single entity, serving a single purpose, which resulted in each stage of the process (input, throughout and output) being focused solely on the needs of the data users. There are obvious advantages to this model of statistics production, in that the individual 'stovepipes' are self-supporting and should a problem arise in one of the stovepipes it usually does not affect other stovepipes.

Three main concerns caused Statistics Netherlands to re-evaluate its 'stovepipe' process. Firstly, the organisation recognised the need for integrated and consistent data, which was not being produced by its current model. Secondly, the administrative costs and burden of the current system was large and the Dutch Government had highlighted as a policy issue a reduction in costs. Thirdly, there was pressure to have the ability to produce the same outputs (statistics) with less staff. Finally, evolution within the information technology sector was creating new possibilities within the statistical methodology sector. Combined, these factors led to a process change within Statistics Netherlands.

Heerschap and Willenborg (2006) highlight that the process involved had a broad scope and a long-term view of 'optimising the statistical production process'. The multifaceted approach taken involved as a first step the identification of strategic corporate goals for Statistics Netherlands. These included:

- *To accommodate the needs of its customers*: This involves a strong emphasis on data integration of individual sets of data on interrelated knowledge bases or theme-oriented bases.
- *To reduce the survey burden of enterprises or customers*: This was to be achieved in two ways – by optimising the use of administrative data and by using consistent terminology and definitions of underlying themes.
- *To develop efficient tools and solutions to statistical issues*.
- *To evaluate existing skills and explore the skills needed to assess the mismatch*. This, in turn, led to adapting the current structure, management and staff to the proposed structure.

The next step involved changing from the 'stovepipe' model to a 'process-driven' model. This required that all separate and duplicate activities were merged. Simply, one central contact centre was set up to deal with incoming requests for data production. This centralised the entire process, but also in effect improved the contactability of Statistics Netherlands as a whole. As Heerschap and Willenborg (2006) state: '*There is a tendency to gain efficiency by merging similar production*

processes that operate as isolated product stovepipes into as few standardised production lines as possible, supported by generic (software) tools in every step of the production chain'.

The core of Statistics Netherlands is the development of its central data repository, based on the dimensions of identified standardised populations, standardised variables and a time dimension. This data repository allows the statistical process to become output-driven, meaning that the process begins with the customer information needs. In essence, Statistics Netherlands is reliant on metadata as a prerequisite – '*Metadata provides the necessary cohesion between different links and tools within the system*'.

As an end goal, Statistics Netherlands strives towards a single standardised production line with a central data repository supported by general tools and metadata and workflow management systems. This is reflected within the change in the architecture of Statistics Netherlands. There are 5 core processes within any one project:

- data collection;
- infrastructure for data storage and data exchange;
- generic tools for processing of data;
- centralised infrastructure for metadata;
- remote access.

Australia

The Australian Bureau of Statistics is working towards building national datasets, with the aim of meeting the statistical requirements of the Australian Government, the State and Territory governments and the community (*see* www.abs.gov.au/). The ABS has identified 11 areas of social concern around which its framework for social statistics is built, namely: population, family and community, education and training, income, culture-leisure, health, housing and neighbourhood, work, crime and justice, employment, transport and travel. In addition, the ABS has identified sub-populations requiring particular attention (Dunlop, 1998). For example, the ABS has produced two data strategies to date – one for children and youth, and the other for cancer.

The information paper, published in 2006, entitled *Improving Statistics on Children and Youth: An Information Development Plan* (Trewin, 2006) had the aim of developing an overview of the national policy and statistical field specifically related to children and youth. This includes identifying relevant Government departments and agencies, major data collections and outputs, as well as main policies, programmes and strategies related to children and youth.

In 2007, Cancer Australia (*see* www.canceraustralia.gov.au/) published the report *A National Cancer Data Strategy*. The aim here was to produce a comprehensive data strategy on cancer information and through a consultation process, to consolidate the data required in key areas by various stakeholders, including researchers, policy-makers, administrators and service providers, as well as the community and their elected officials. Then an identification of all of the available data sources was undertaken, with information on what types of data are provided and who provides the data. In developing this data strategy, emphasis was placed on 4 key action areas:

- *To improve data availability*, which involved the following activities: (1) assess and prioritise data needs for cancer control; (2) map data availability in relation to these needs; (3) identify and prioritise gaps and collaborate with data collection agencies in closing gaps; (4) address deficiencies in indigenous identification in cancer registries and allied databases; (5) ensure that data use is maximised; (6) removal of any excess barriers to data usage; and (7) promote awareness among funding bodies of data applications in cancer control and ensure that key data collection activities are continued.
- *To improve data reporting*, primarily involving the promotion of the production of data reports that address key cancer questions.
- *To improve data quality and consistency*.
- *To promote research in data collection and use*.

USA

The US Census Bureau (*see* www.census.gov/) utilises a statistical metadata repository, with the goal of building '*a metadata repository with Internet/intranet access to support Internet data dissemination and automated survey design and processing tools*' (Gillman *et al*, 1998). It is intended that the repository contains information that describes the design, processing, analysis and data for all of the surveys conducted by the Census Bureau. Metadata is collected automatically as a function of the development of survey design and processing tools.

The US Census Bureau outlines 3 models of metadata which are required to meet its needs. Together, these models describe the underlying variables that are measured or reported by these surveys, although the authors Gillman and Appel (2000) recognise that the process is laborious and not easy. The first two models – a survey process model and a business data model – were developed to subsequently describe as well as document the design, processing and analysis of statistical surveys. The third model – the data element model – outlines the information about the variables included with the surveys.

Work is underway to further develop the current statistical metadata repository towards a corporate metadata repository. This repository has been established to support the statistical information system (SIS). The design of the corporate metadata repository is based on 3 data models, which are integrated into one extensive model within the repository (Gillman and Appel, 2000). The first model utilised is the *business data model*, which is primarily concerned with survey designs, processing, analyses, datasets, products and documents as related to the surveys. The second model is the *data element model*, which outlines the structure for managing the names, definitions, permissible values and other attributes of data elements. The final model is the *metamodel*, which describes issues such as security, access control, database schemas, record layouts and timeframes.

Finally, the overall model is subdivided into 5 functional areas (data element, registration, metamodel, business data and documentation), which outline how the model works (Gillman *et al*, 1998).

References

Adelman, S. (2001) 'Data Strategy Introduction', *DM Review*. Available at: www.dmreview.com/dmdirect/20011116/4291-1.html

Agosta, L. (2007) 'Data Warehousing Trends for 2007: Data Strategy Advisor', *DM Review*. Available at: www.dmreview.com/news/1069307-1.html

Anderson, R., Brown, I., Clayton, R., Dowty, T., Korff, D. and Munro, E. (2006) *Children's Databases – Safety and Privacy: A Report for the Information Commissioner*. Wilmslow, UK: Foundation for Information Policy Research.

Article 29 Data Protection Working Party (2007) *Working document on the processing of personal data relating to health in electronic health records (EHR)*. Brussels: The European Commission. Available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf

Ballou, D.P. and Tayi, G.K. (1999) 'Enhancing data quality in data warehouse environments', *Communications of the ACM*, Vol. 42, No. 1, pp. 73-78.

Blobel, B. (2004) 'Authorisation and access control for electronic health records systems', *Journal of Medical Informatics*, Vol. 73, pp. 251-57.

Budgen, D., Rigby, M., Brereton, P. and Turner, M. (2007) 'A data integration broker for healthcare systems', *IEEE Computer Society*, Vol. 40, No. 4, pp. 34-41.

Cancer Australia (2007) *A National Cancer Data Strategy for Australia: A collaborative approach to improving cancer outcomes through cancer data*. Canberra: Australian Government. Available at: http://canceraustralia.gov.au/sites/default/files/user-upload/publications/ncds_final_web.pdf

Carson, C.S. (2001) *Toward a framework for assessing data quality, IMF Working Paper No. 01/25*. Washington, DC: Social Science Research Network.

CIHI (2006) *Earning Trust – 3 Years Later*. Ottawa: Canadian Institute for Health Information.

CIHI (2004) *What is a unique identifier?* Ottawa: Canadian Institute for Health Information. Available at: http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=infostand_uniquemore_e

CIHI (2003) *Earning Trust: Key Findings and Proposed Action Plan from the Data Quality Strategies Study*. Ottawa: Canadian Institute for Health Information.

CIHI (2001) *The CIHI Data Quality Framework (Revised)*. Ottawa: Canadian Institute for Health Information. Available at: http://secure.cihi.ca/cihiweb/en/downloads/Data_Quality_Framework_2004_e.pdf

CNIL (2001) *The Internet and the collection of personal data from minors*. Paris: National Commission for Informatics and Freedoms. Available at: www.cnil.fr/english/

Dao, S. and Perry, B. (1996) 'An overview of data mining in heterogeneous schema integration', *IEEE Computer Society*, Vol. 3, pp. 478-83.

Department of Health and Children (2004) *Health Information: A National Strategy*. Dublin: Government Publications. Available at: www.drugsandalcohol.ie/5862/

Dunlop, B. (1998) *Building National Data Sets – An ABS Perspective*, Presentation to the CATI Population Health Survey Forum, Melbourne, Australia, October 1998. Available at: www.dhs.vic.gov.au/nphp/catitrg/forum98/bd.htm

Elliot, H. and Popay, J. (2000) 'How are policy-makers using evidence? Models of research utilization of local NHS policy-making', *Journal of Epidemiology and Community Health*, Vol. 54, pp. 461-68.

European Standards on Confidentiality (2005) Available at: www.eurosocap.org

Fikes, R., Farquhar, A. and Pratt, W. (1996) *Information Brokers: Gathering information from heterogeneous information sources*, Paper presented to the Ninth Florida Artificial Intelligence Research Symposium, Florida, USA, Spring 1996.

Gillman, D.W. and Appel, M.V. (2000) *Statistical metadata research at the Census Bureau.* Washington, DC: US Census Bureau.

Gillman, D.W., Appel, M.V. and Highsmith, S.N. (1998) *Building a statistical metadata repository at the US Bureau of the Census*, Paper presented at the Statistical Commission and Economic Commission for Europe Conference of European Statisticians, Geneva, Switzerland, February 1998.

Grossmann, W. (2004) *Metadata models in survey computing*, Paper presented at the meeting of Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS), Geneva, Switzerland, February 2004.

Grossmann, W. (1997) *Use of metadata in the statistical production process.* Vienna: University of Vienna.
Hackett, Y. (2001) 'A national research data management strategy for Canada: The work of the National Data Archive Consultation Working Group', *IASSIST*, Vol. 25, No. 3, pp. 13-16.
Harries, U., Elliot, H. and Higgins, A. (1999) 'Evidence-based policy-making in the NHS: Exploring the interface between research and the commissioning process', *Journal of Public Health Medicine*, Vol. 21, No. 1, pp. 29-36.
Haynes, C.L., Cook, G.A. and Jones, M.A. (2007) 'Legal and ethical considerations in processing patient-identifiable data without patient consent: Lessons learnt from developing a disease register', *Journal of Medical Ethics*, Vol. 33, pp. 302-7.
Health Information Strategy Steering Committee, New Zealand (2005) *Health Information Strategy for New Zealand 2005*. Wellington, NZ: Ministry of Health. Available at: www.moh.govt.nz/moh.nsf/0/edb64619d460f974cc2570430010fb71?OpenDocument
Heerschap, N. and Willenborg, L. (2006) 'Towards and integrated statistical systems at Statistics Netherlands', *International Statistical Review*, Vol. 74, No. 3, pp. 357-78.
Hull, R. (1997) 'Managing semantic heterogeneity in databases: A theoretical perspective', *Proceedings of the Sixteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of database systems*. Tuscon, AZ: ACM Press, pp. 51-61.
Inmon, W.H. (1996) *Building the Data Warehouse* (2nd edition). New York: John Wiley.
INIsPHO (2007) *Metadata standards for Ireland and Northern Ireland's Population Health Observatory (INIsPHO) and All-Ireland electronic Health Library (AIeHL), Version 2.0*. Dublin: Institute of Public Health. Available at: www.publichealth.ie/files/file/MetadataStandards_for_INIsPHOandAIeHL_V2_0%5B1%5D.pdf
IPSMS (2007) *Irish Public Service Metadata Standard, Version 1.0. Part 1: Framework*. Dublin: Irish Public Service Metadata Standard. Available at: www.gov.ie/webstandards/metastandards/ipsms_part1.pdf
Iversen, A., Liddell, K., Fear, N., Hotopf, M. and Wessely, S (2006) 'Consent, confidentiality and the Data Protection Act', *British Medical Journal*, Vol. 332, pp. 165-69.
Iezzoni, L.I. (1997) 'Assessing quality using administrative data', *Annals of Internal Medicine*, Vol. 8, No. 2, pp. 666-74.
Jeskanen-Sundström, H. (2007) 'Needs for change and adjusting them in the management of statistical systems', *Statistical Journal of the IAOS: Journal of the International Association for Official Statistics*, Vol. 24, pp. 85-91.
Kavanagh, P., Balanda, K.P. and Shortt, N. (2005) *Metadata standards for Ireland and Northern Ireland's Population Health Observatory, Version 1.0*. Dublin: Institute of Public Health.
Kimball, R., Reeves, L., Ross, M. and Thornwaite, W. (1998) *The Data Warehouse Lifecycle Toolkit*. New York: Wiley.
Kerr, K. (2002) *The development of a data quality framework and strategy for the New Zealand Ministry of Health*. Auckland: University of Auckland.
Klein, B. and Rossin, D.F. (1999) 'Data errors in neural network and linear regression models: An experimental comparison', *Data Quality*, Vol. 5, No. 1, p. 25.
Korin, J.B. and Quattrone, M.S. (2007) 'Electronic Health Records raise new risk of malpractice liability', *Law Technology News*. Available at: www.law.com/jsp/lawtechnologynews/PubArticleFriendlyLTN.jsp?id=900005483988&slreturn=1
Lenz, H.J. (1994) *The conceptual schema and external schemata of metadatabases*, Paper presented at the Seventh International Working Conference on Scientific and Statistical Database Management, Virginia, USA, September 1994.
Liaw, T., Lawrence, M. and Rendell, J. (1996) 'The effect of a computer-generated patient-held medical record summary and/or a written health record on patients' attitudes, knowledge and behaviour regarding health promotion', *Journal of Family Practice*, Vol. 12, pp. 289-93.
Lawlor, D.A. and Stone, T. (2001) 'Public health and data protection: An inevitable collision or potential for a meeting of minds?', *International Journal of Epidemiology*, Vol. 30, pp. 1221-25.
McClelland, R. and Harper, C.M. (2007) European guidance for healthcare professionals on confidentiality and privacy in health care. Belfast: Queen's University, Belfast. Available at: www.qub.ac.uk/methics/HarperMcClelland.pdf
McDonagh, M. (2006) *Freedom of Information Law*. Dublin: Thomson Roundhall.
NSB (2004) *Best Practice Guidelines for the development and implementation of formal data/statistics strategies in Government Departments*. Dublin: National Statistics Board. Available at: www.nsb.ie/pdf_docs/Data_Strategy_Guidelines.pdf
Office of the Data Protection Commissioner (2008) *Data protection guidelines on research in the health sector.* Dublin: Office of the Data Protection Commissioner. Available at: www.dataprotection.ie/documents/guidance/Health_research.pdf

Oracle (2006) *Oracle Customer Case Study: Australian Bureau of Statistics gains powerful analytic facilities with integrated data warehouse*. Available at: www.oracle.com/customers/snapshots/australian-bureau-of-statistics-dw-linux-casestudy.pdf
Redman, T.C. (2001) *Data Quality: The Field Guide*. Boston: Digital Press.
Rigby, M.J., Budgen, D., Brereton, O.P., Bennett, K., Layzell, P., Keane, J., Russell, M., Kotsiopoulos, I., Turner, M. and Zhu, F. (2005) 'Proving the concept of a data broker as an emergent alternative to supra-enterprise EPR systems', *Medical Informatics and the Internet in Medicine*, Vol. 30, No. 2, pp. 99-106.
Rigby, M., Budgen, D., Turner, M., Kotsiooulos, I., Brereton, P., Keane, J., Bennett, K., Russell, M., Layzell, P. and Zhu, F. (2007) 'A data-gathering broker as a future orientated approach to supporting EPR users', *International Journal of Medical Informatics*, Vol. 76, pp. 137-44.
Schenk, K., Murove, T. and Williamson, J. (2006) 'Protecting children's rights in the collection of health and welfare data', *Health Human Rights*, Vol. 9, No. 1, pp. 80-100.
Siltala, H. (2005) *Inside SOTKAnet.* Available at: www.heikkisltala.com/SOTKAnet.htm
Srivastava, J. and Chen, P.Y. (1999) 'Warehouse creation: A potential roadblock to data warehousing', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 11, No. 1, pp 118-26.
Staroselsky, M., Volk, L.A., Tsurikova, R., Pizziferri, L., Lippincott, M., Wald, J. and Bates, D. (2006) 'Improving electronic health record (EHR) accuracy and increasing compliance with health maintenance clinical guidelines through patient access input', *Informational Journal of Medical Informatics*, Vol. 75, pp. 693-700.
Strong, D.M., Lee, Y.W. and Wang, R.Y. (1997) 'Data quality in context', *ACM*, Vol. 40, No. 5, pp. 103-10.
Sundgren, B. (1996) 'Making statistics data more available', *International Statistical Review*, Vol. 64, No. 1, pp. 23-38.
Sundgren, B. (1993) *Guidelines on the design and implementation of statistical metainformation systems*, Paper presented at the Conference of European Statisticians, Geneva, Switzerland, February 1993.
Trewin, D. (2006) *Improving Statistics on Children and Youth: An Information Development Plan*. Canberra: Australian Bureau of Statistics.
Wand, Y. and Wang, R.Y. (1996) 'Anchoring data quality dimensions in ontological foundations', *Communications of the ACM*, Vol. 39, No. 11, pp. 86-95.
Wang, R.Y., Kon, H.B. and Madnick, S.E. (1993) 'Data quality requirements analysis and modeling', *IEEE Computer Society*, Vol. 4, No. 4, pp. 670-77.
Wang, R.Y., Storey, V.C. and Firth, C.P. (1995) 'A framework for analysis of data quality research', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 7, No. 4, pp. 623-40.
Widom, J. (1995) *Research problems in data warehousing*, Paper presented at the Conference on Information and Knowledge Management, Maryland, USA, July 1995.
UN/ECE (1995) *Guidelines for the modeling of statistical data and metadata.* United Nations Statistical Commission and Economic Commission for Europe. Available at: www.unece.org/stats/publications/metadatamodeling.pdf

Websites

Australian Bureau of Statistics – **www.abs.gov.au/**

Data Protection Commissioner (Ireland) – **www.dataprotection.ie/docs/Home/4.htm**

European Standards on Confidentiality and Privacy (EuroSOCAP) – **www.eurosocap.org/**

Ireland and Northern Ireland's Population Health Observatory – **www.inispho.org/**

Statistics Netherlands – **www.cbs.nl/en-GB/menu/home/default.htm**

Universal Declaration on Bioethics and Human Rights (2005) – **www.unesco.org/new/en/social-and-human-sciences/themes/bioethics/bioethics-and-human-rights**

UN Convention of the Rights of the Child (1989) – **http://www2.ohchr.org/english/law/crc.htm**

UN Statistical Commission (2007) – **http://unstats.un.org/unsd/default.htm**

US Census Bureau – **www.census.gov/**

US Environmental Protection Agency, *Children's Health Protection: Scientific Data and Methods* – **http://yosemite.epa.gov/OCHP/OCHPWEB.nsf/content/Whatwe_scientif.htm**

Statutory Instruments

Ireland
Data Protection (Access Modification) (Health) Regulations 1989 (SI No. 82 of 1989)

Data Protection (Access Modification) (Social Work) Regulations 1989 (SI No. 83 of 1989)

Case Law

Ireland
Kennedy v Ireland [1987] IR 587
McGee v Attorney General [1974] IR 284

United Kingdom
Gillick v West Norfolk and Wisbech Area Health Authority and the Department of Health and Social Security [1985] UKHL 7

Europe
Amann v Switzerland [GC], No. 27798/95
Census judgment, German Constitutional Court (1983)
Gaskin v United Kingdom (1989) 12 EHRR 36
Guerra v Italy (1998) 26 EHRR 357
Leander v Sweden (1987) 9 EHRR 433
Sunday Times v United Kingdom (1979-80) 2 EHRR 245