The Clinical Data Repository Provides CPR’s Foundation

The core of any computer-based patient record system is a permanent data store. The CDR core capability will evolve as CPR systems mature.

The provision of clinical care, particularly in an inpatient hospital environment, is a complex process involving the interaction of many caregivers (physicians, nurses, technicians, staff) to carry out a wide variety of processes (such as admit a patient, administer a course of antibiotics, perform a surgical procedure) using a comprehensive and diverse set of data. Central to successful clinical automation support for these processes is the underlying data store — the clinical data repository (CDR), a core capability of the enterprise computer-based patient record (CPR) system.

The other eight core capabilities as defined by Gartner are:

- Clinical documentation and data capture (see "Clinical Documentation in the Enterprise CPR")
- Clinical workflow (WF; see "Workflow Evolution Within the CPR Generations")
- Clinical display — including the clinical dashboard (see "The Clinical Dashboard Is Another Key CPR System Element")
- Order management — including physician order entry (see "Order Management Is a Core Capability of a CPR System")
- Clinical decision support (CDS; see "Clinical Decision Support: Core Capability of Evolving CPR")
- Support for privacy
- Interoperation (see "The Gartner 2004 Criteria for the Enterprise CPR")
- Controlled medical vocabulary (CMV; "Controlled Medical Vocabulary in the CPR Generations")

At its core, a CPR system must have a permanent data store that can deal with data from a variety of sources organized in a patient-centric manner. This is accomplished by the CDR, which
is a fundamental building block of the CPR system. The CDR must make patient data available to the other CPR system core capabilities in a manner that is sufficiently timely to support patient care. It must track amendments to the data in a matter that permits retrieving the information as it was before and after an amendment. (This is no different than the discipline for amendments that is imposed on those who make entries in the paper medical record.)

Because CDR data is available more rapidly than data in the paper chart, and because it can be viewed simultaneously in multiple locations, providers have the opportunity to make care decisions based on information that is in the CDR but not yet in the paper chart. Institutions with advanced CPR systems will rely on the paper chart less and less; advanced users of the CPR system ultimately will go "paperless" and rely on the data from the CDR all of the time.

Because the CDR data is available more widely than the paper chart, the CPR system requires careful attention to the authorization and tracking of access to patient data. This privacy protection can be facilitated by design features of the CDR.

Reliance on the CDR for care processes implies that the it must store data in a manner that supports business continuity despite failures attributable to hardware, software or the environment. For many clinical environments, the CDR data must be available around the clock, everyday. This reduces the techniques that a vendor may apply for business continuity. For example, offline backups are not a useful approach.

The CDR must be available to store and use data in a multiplicity of forms including textual reports stored as document images, textual reports stored as text — or Extensible Markup Language (XML), diagnostic and photographic images, as well as finely structured, explicitly coded data amenable to computer processing for decision support, smart retrieval and other highly automated functions. It also must support structured, cross-patient analysis of the data for research, quality, patient safety and other purposes. Such analysis must be carefully controlled to avoid violations of patient privacy, and may require a formal mechanism to deidentify patient information.

Gartner believes that as a practical matter such a data store should be at least as flexible and capable of rich expression as the Health Level 7 Reference Information Model (HL7 RIM). If it is not, vendors will have trouble importing rich clinical data from other sources or keeping up with the expansion of capability likely to come from vendors with richer information models. Also, if the data in the CDR is to be maximally useful, it is essential
that a there be an enterprisewide standard for the medical concepts, terms and relationships that are used — a CMV.

At the same time, it is important to recognize that CPR systems will continue to receive data in images and textual reports for many years. The challenge to the CPR system is to represent the mixture of image and text data, as well as finely structure coded data, in a way that gleans the most value from each.

Many CPR systems use different storage mechanisms for images and character data. This approach requires special care when dealing with data recovery for business continuity or archival recovery.

The increasing reliance on data from the CDR engenders additional requirements for archival storage of CDR data. The data must be available immediately, when it can be reasonably anticipated that the patient will be treated, and it must be available online or from archival storage for the legally required time frame. (That time frame differs by state, but typically ranges from seven to 25 or more years). This is true even if the data is never actually needed or accessed during that time frame. The capability to view data as it was before and after amendments becomes critical when it is being retrieved for forensic capabilities.

Being able to view archival data has implications for the entire CPR system. It is insufficient to be able to simply retrieve data, it must also be possible to view it as it originally appeared. This implies that the CPR system vendor never make an image or text format obsolete through software revisions or make incompatible extensions to the database design. The vendor must maintain the historical values of codes, even when the organizations that promulgate the code may decide to re-use the same code value for different purposes. Ideally, a CPR system would record or be able to reconstruct how archival data appeared when it was viewed by a provider.

Today, many CPR system vendors are treating the requirements for business continuity and archival recovery as if they can be fully handled by storage area networks or other technologies external to the CPR system. As data volume and record retention times grow, particularly as image data becomes part of the logical electronic record, there is likely to be an interaction between the architecture of a CPR system and business continuity, and records retirement and recovery approaches. Care delivery organizations (CDOs) acquiring CPR systems should evaluate this issue based on five-year projections of data storage requirements.
As the amount of data in the CPR system increases, not only can performance suffer, but backups and disaster recovery also can be adversely affected. Therefore, CPR systems must include archival capabilities, something that most available products do not adequately address. Other issues that must be addressed include fault tolerance, disaster recovery, schema extensions, concurrent updates and archival maintenance.

**CDR Capabilities Across the CPR System Generations**

Here is a description of the CDR capabilities required to achieve various levels of Gartner CPR system generation functionality (currently, most CPR systems are at Generation 2, but several Generation 3 systems are expected to be available by year-end 2004):

- **Generation 1, the Collector:** These have simple database designs with minimal, if any, structured data. The approach to business continuity may involve substantial offline time when recovering from a disaster. There is no formal means for offloading data to a long-term archive. The CDR itself provides little or no capability to protect patient privacy. Provisions for cross-patient reporting are not required.

- **Generation 2, the Documentor:** At this level, the CDR must have a uniform file system with a flexible underlying data structure that can organize and manipulate a mixture of fine-grained clinical information and summary text. Once "committed," information cannot be changed or deleted, only annotated or amended. This generation still has no formal means for offloading data to a long-term archive, and provisions for business continuity continue to require downtime for recovery. If cross-patient reporting capabilities exist, they are available only to a few analysts who are responsible for ensuring that reports respect the institutional rules protecting patient privacy.

- **Generation 3, the Helper:** This requires a database management system and independent metadata, as well as a built-in means for referential integrity based on logic rules. Medical data must be represented in a variety of forms (for example, textual reports stored as document images, textual reports stored as text or XML), and finely structured, explicitly coded data amenable to computer processing. The CDR database design supports the creation of rules that can restrict provider access to patient data based on facility-specific rules (as well as a "break the glass" mechanism to override these provisions). The data store must be at least as flexible and capable of rich expression as HL7 RIM. Recovery from many hardware and software failures is automatic and completed in a few minutes. There is some simple support (such as "download data for any patient who
has not been seen for five years") for automatically selecting data for archiving and removal from the active database with a mechanism for retrieving data as needed from near-line or offline storage.

• Generation 4, the Colleague: Structured, cross-patient analysis of data with appropriate mechanisms to protect patient privacy is required. The systems must have the capability to process incoming and outgoing clinical information to permit full, fine-grained exchange with any other intraenterprise automated facilities. There must be sophisticated archiving capabilities that maximize performance while continuing to guarantee access to necessary data and providing flexibility for sites to create their own rules for offloading data to the long-term archive. The systems also make it possible to reconstruct how archived data was originally presented to clinicians.

• Generation 5, the Mentor: There must be multilingual support for all languages used in a CDO. The system permits full, fine-grained exchange of clinical information with automated facilities across enterprises.

Key Facts:

• Access to relevant clinical data is essential if healthcare delivery is to be efficient and effective.

• Successful clinical automation requires a CDR that maximizes performance while guaranteeing access to the data.

• As CPRs mature and the amount of data they contain grows, an effective method for offloading data to a long-term archive will be essential.

Bottom Line: Automated support for the processes of care delivery is a foundational requirement for care delivery organizations that are attempting to support the practice of medicine, eliminate errors, improve quality, reduce unnecessary practice variation and control costs. If a computer-based patient record system is to be effective, the underlying clinical data repository must be able to guarantee performance, as well as data integrity and reliable access to data. Areas of weakness in current enterprise CPR systems include dealing with structured data, provisions for business continuity and long-term archival capabilities. CDOs considering the purchase of a CPR system must ensure that their vendors have adequate CDR functionality now and clear strategies to continue enhancing these capabilities as their system evolve.

Acronym Key

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDO</td>
<td>care delivery organization</td>
</tr>
<tr>
<td>CDR</td>
<td>clinical data repository</td>
</tr>
<tr>
<td>CMV</td>
<td>controlled medical vocabulary</td>
</tr>
<tr>
<td>CPR</td>
<td>computer-based patient record</td>
</tr>
<tr>
<td>HL7 RIM</td>
<td>Health Level 7 Reference Information Model</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
</tbody>
</table>