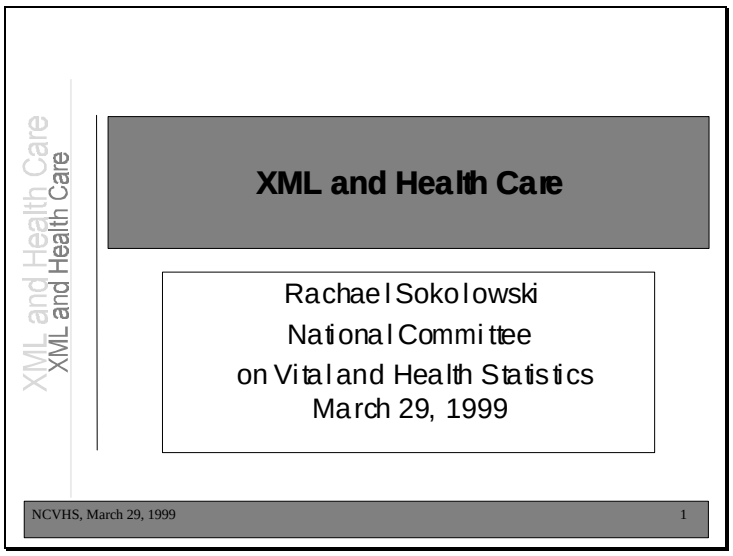
Slide 1



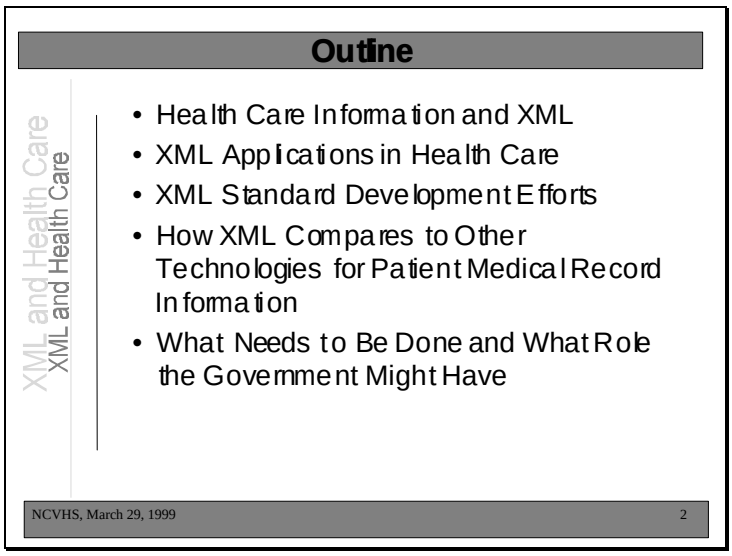
Rachael Sokolowski
rsokolowski@magnoliatech.com
rsokolowski@itrust.net

Primary affiliation:
President, Magnolia Technologies
Consulting company specializing in SGML and XML for health care and health care publishing.

Secondary affiliation:
Vice President of Research, iTrust
Software development organization, SGML electronic health record and practice management system.

Standards Development Organizations
Co-chair, Health Level 7, HL7 SGML/XML Special Interest Group
Co-chair, ASTM E31.25, "XML DTDs for Health Care"

Slide 2



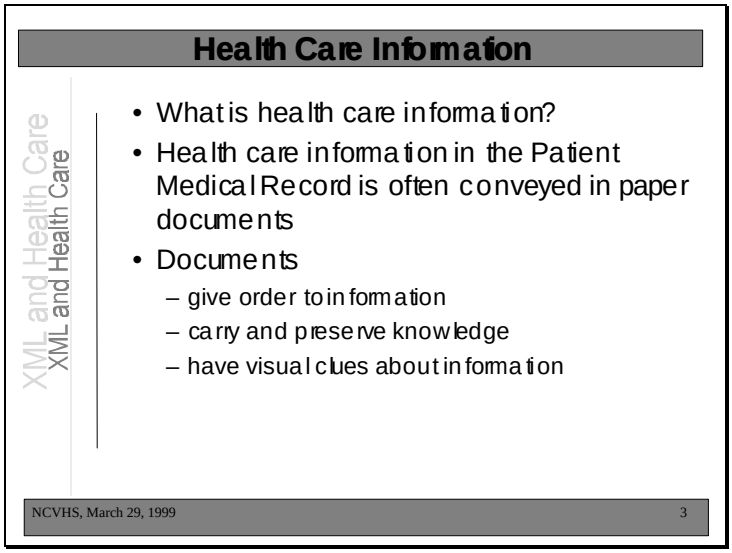
Introduction

Recently there has been widespread and increased use of the World Wide Web by the health care industry. The use of Web technologies has not removed the barriers to accessing and exchanging health care information contained in a Patient's Medical Record (PMR). However, there are new opportunities for the representation and exchange of clinical information emerging from the convergence of data processing, communication, and publishing technology, in a new standard for the Web, eXtensible Markup Language (XML).

In the past two years, a growing number of providers, vendors, and standards organizations have begun investigating how XML, a simplified subset of SGML (Standard Generalized Markup Language) designed for Web delivery, can solve problems in health care information systems. XML is a platform, vendor, and application independent technology for describing a document's content and structure.

There are a number of things that must be done for a successful implementation of XML for Patient Medical record Information. This presentation outlines some of these needs.

Slide 3

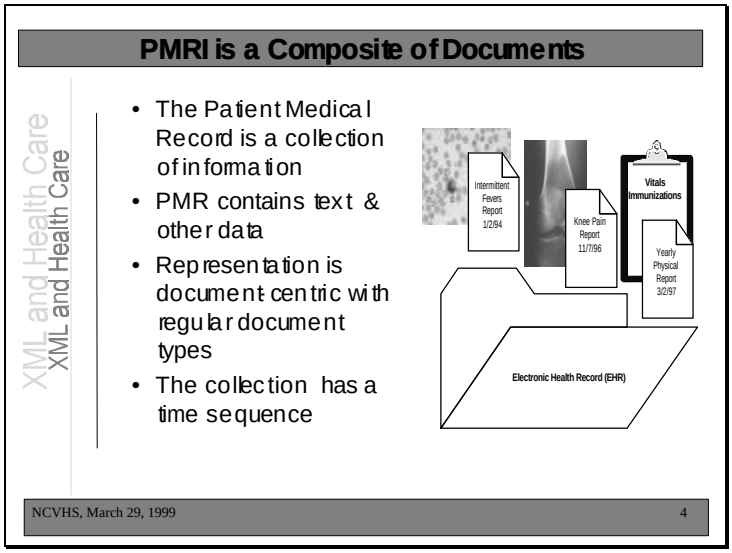


No industry stands to benefit more from electronic information than the health care industry. Each year health care information technology becomes more complex and heterogeneous reflecting the dynamic nature of the industry.

The information processing needs of individual health care providers pose increasing challenges for technology solutions. The quality of information processing ultimately effects the quality of patient care itself. Health care has an additional problem when exchanging and processing information: the difficulty in reducing the physician or caregiver's notes to regular, predictable, and discrete data points. That can be kept in a PMR.

Human language, narrative, has been central to medical records. Most of the work on computer-based patient medical records systems has assumed that free form narrative must be replaced by data so that the crucial information contained in the record can be processed by a computer. Until recently, "narrative" and "machine-readable" were mutually exclusive. Electronic documents offer the ability to represent information a form that is human readable and can be processed by a computer.

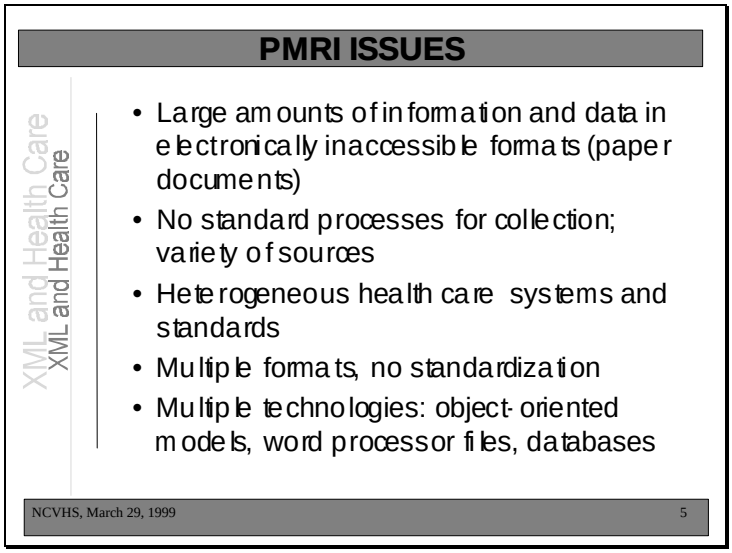
Slide 4



The Patient Medical Record is a collection of information. It is typically a folder of paper documents, collected over time and placed in the folder. Other information outside the folder is routinely used by health care professionals to keep notes, lists of allergies or current medications or chronic problems.

The electronic is a computer-based Patient Medical Record may be generated by many sources such as transcription, scanned from paper, created by and structured reporting software systems. The electronic document may be rendered in many forms such as printed to paper or a computer screen, stored a tables in a database or transformed into other software representations such as messages or transactions.

Slide 5

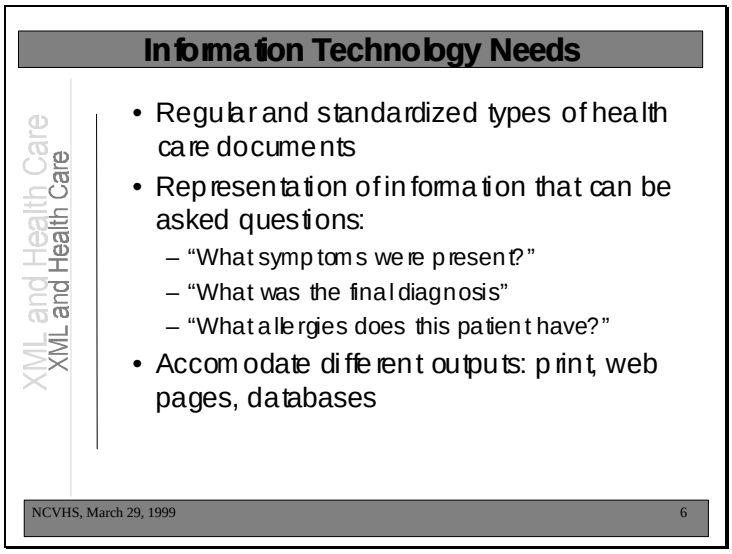


Information about patients is critical for delivery of high quality service, but this information exists in representations that make access difficult or impossible. A single point of care, such as a hospital or clinic, may draw patient information from laboratory systems, scheduling and billing systems, imaging systems, and many sources of paper records including records of care from other sites and organizations.

The clinical patient record has been entirely focused on being in a form that is human readable and has ignored the requirements for machine processing. health care information standards, such as HL7 (Health Level 7) and EDI X12 (Electronic Data Interchange), have been almost entirely focused on machine interaction and have ignored the paper-based record. The slow emergence of computer based patient records has led to a classification of information gathered during clinical encounters. This classification is needed as part of the patient record and as a resource for health research and statistics. The classification and coding requirements of these two purposes differ, patient medical records require as much specific detail as possible, where as health statistics require data which are systematically aggregated into categories based on their frequency or their importance for policy.

We need to reach a middle ground. Documents need to be in a form in which humans are able to read and analyze and one in which machines are able to process. XML has the potential to reach this middle ground.

Slide 6

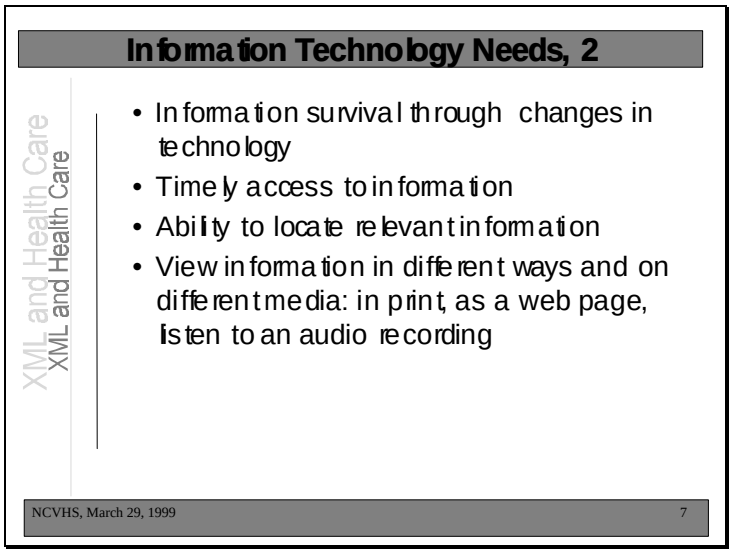


There are some easily identifiable requirements for health care information. A set of the types of electronic documents in a PMR needs to be derived. Electronic documents need a standard structural representation of the content so that questions of the information can be asked, and answers easily found. For example:

- Collection and Access: "What was the diagnosis?"
- Quantification: "How many physical exams were performed?"
- Narrative Text: Retain observer's notation

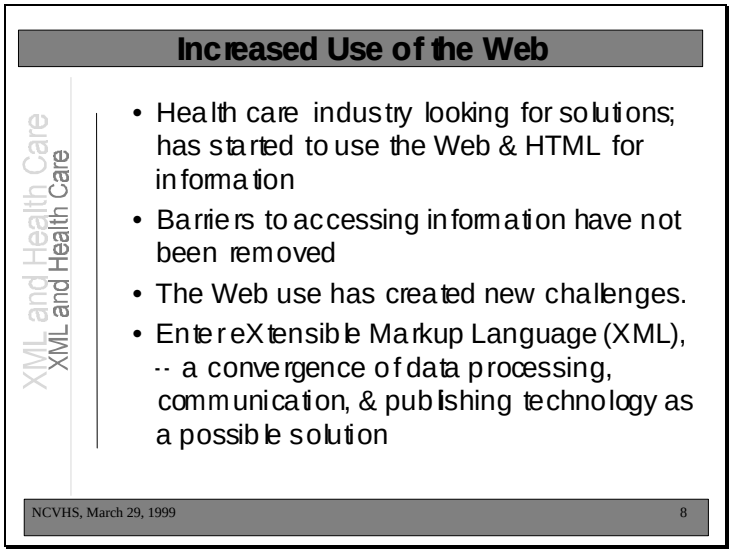
The information must be published on various different outputs.

Slide 7



Patient Medical Record information needs to be in a form that will exist for the lifetime of a patient and in a form that is likely to survive changes in technology. Health care professionals want to have access to relevant information quickly and when it is needed. Lastly and most importantly, information needs to be correct and be from trusted sources.

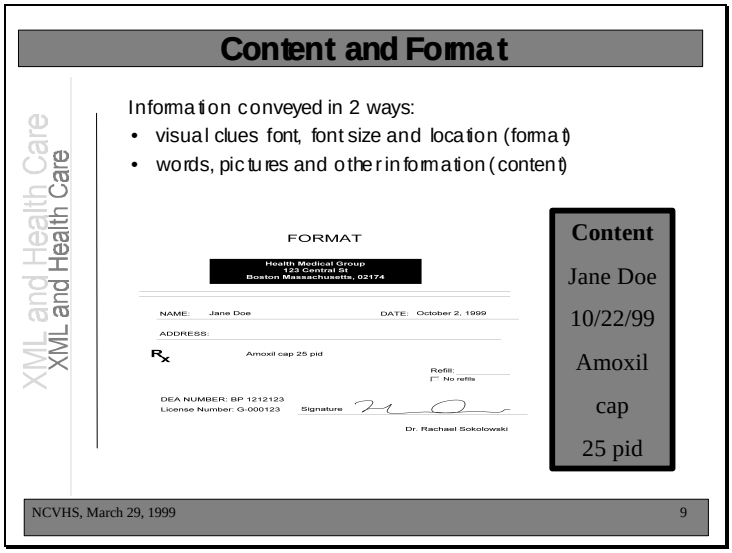
Slide 8



The vast majority of data exchange in health care is still with proprietary technologies, but there is a rapidly emerging interest in using the Web, and web-based technologies for this purpose. These web technologies are a typically combination of proprietary programmatic interfaces to manipulate the content, structure and style of HTML documents (which are also known as web pages).

HTML was created to describe and display pages of text on a computer screen connected to the World Wide Web. HTML describes how to display the data but does not offer any information about the content of the information so it is an insufficient representation for data generated in clinical settings. Web technologies are a convenient and low cost data capture and distribution mechanism for organizations that are geographically and functionally diverse and are gaining wide use in the health care industry for these reasons. Software components that translate information between HTML, databases and EDI are widely used, but are one-off, proprietary solutions that do not enable widespread information transfer.

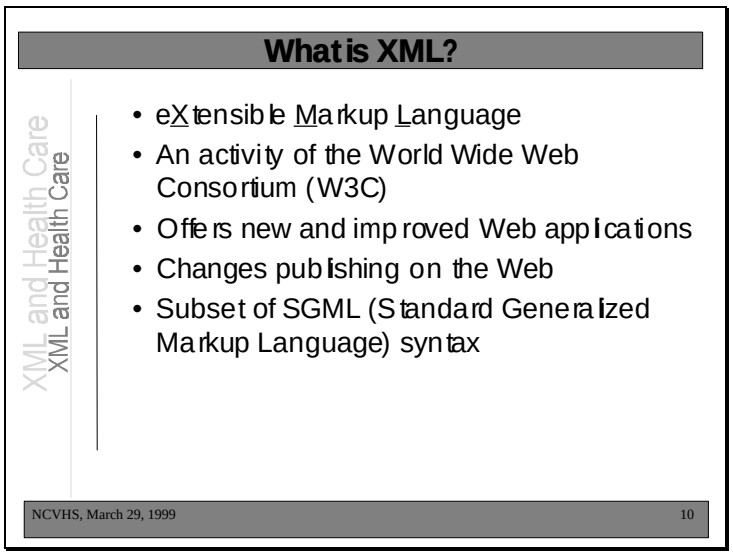
Slide 9



It is not known if health care documents have common and identifiable structures but it is generally accepted that some regularity exists. This assumption was based on analyzing printed documents. Printed documents have information conveyed in two ways: the content of the document and the format of the content. The content contains the words, pictures and other information and the format provides visual clues font, font size and location.

Items that are presented in a style different from the paragraph text of a document are likely to have some semantic meaning.

Slide 10



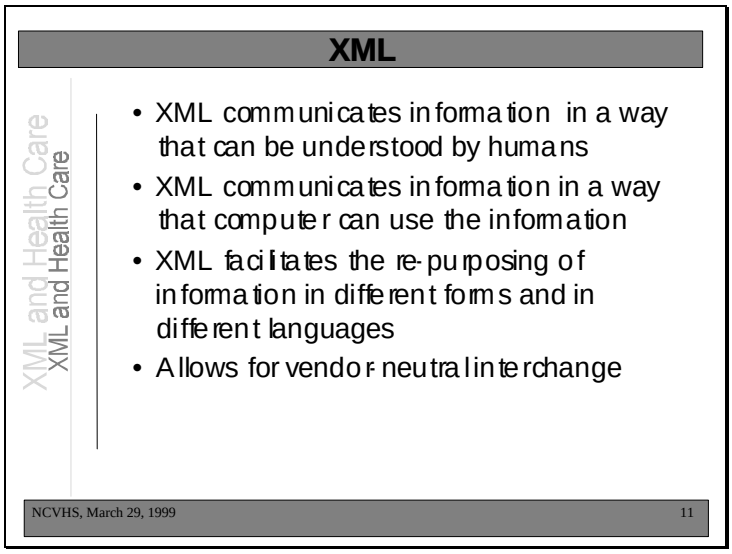
The past ten years have seen growing use of text processing markup languages outside of health care that make possible the rigorous specification, processing, and analysis of information in narrative form. These markup languages are based on Standard Generalized Markup Language (SGML). SGML is a meta-language, that is, a language that sets up rules for creating markup languages called applications. HyperText Markup Language (HTML) is an application of SGML.

Recently, a simplified subset of SGML has been specified called eXtensible Markup Language or XML.

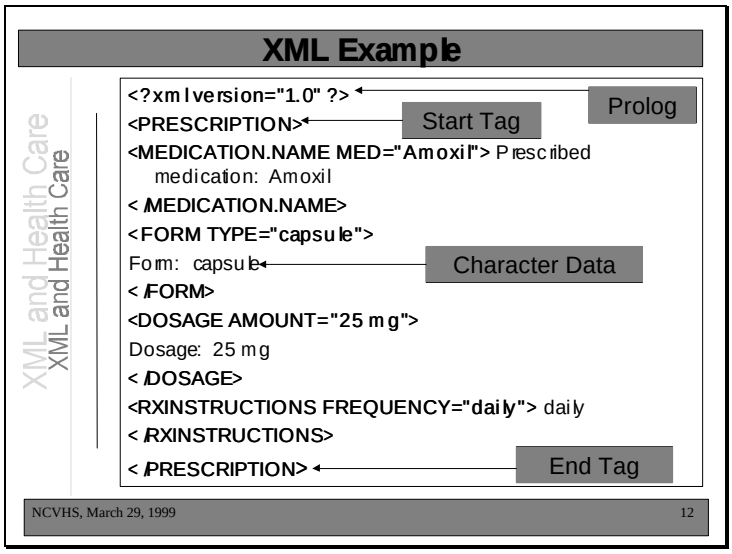
In health care, SGML and XML present great promise for allowing exchange and processing of clinical information without sacrificing the precision, nuance, and complexity of human language. While the feasibility of SGML and XML for clinical records is gaining widespread acceptance, this has not been tested in live, clinical contexts supporting multi-faceted exchange of data, text, pictures, and other media.

Reference: Light, Richard, Presenting XML, Sams.net, Indianapolis, IN, 1997, p.343

Slide 11



Slide 12



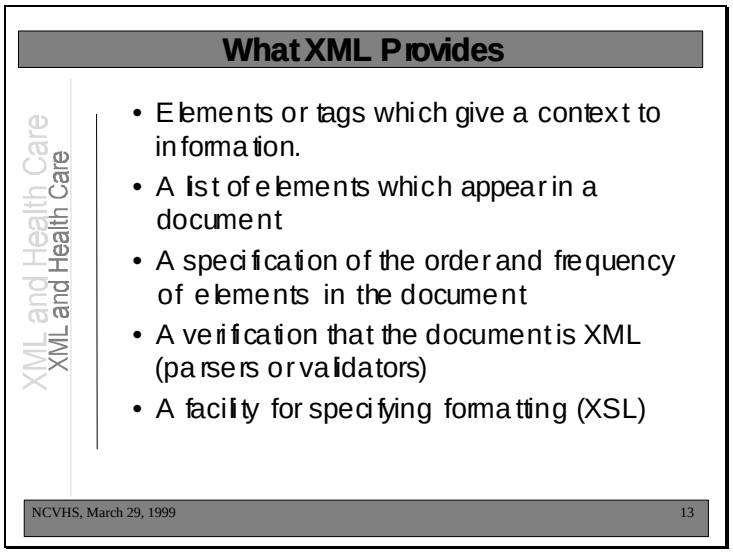
XML documents identify themselves as XML with the fist line of the document:
<?xml version="1.0" ?>

XML adds tags which provide further information about the content of the document. In health care , important or meaningful data can be easily identified by XML markup or tags such as <SUBJECTIVE> and <DIAGNOSIS>. Tags are a used to mark the structure of an XML document and to provide context for information. Notice in the above example it can be determined that the family name given here is for a patient and not a provider.

The tags are not “seen” in the published version of the electronic document. In the example above the following could appear in the document:

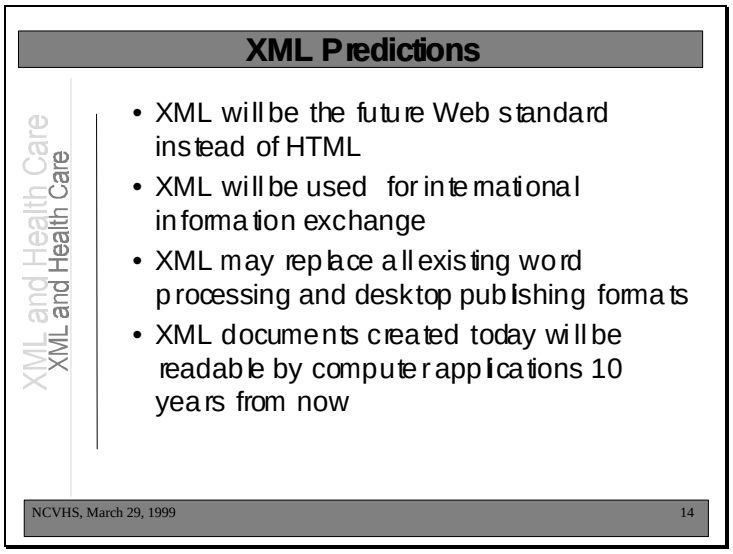
Prescribed medication: Amoxil **Form:** capsule **Dosage:** 25 mg. daily

Slide 13



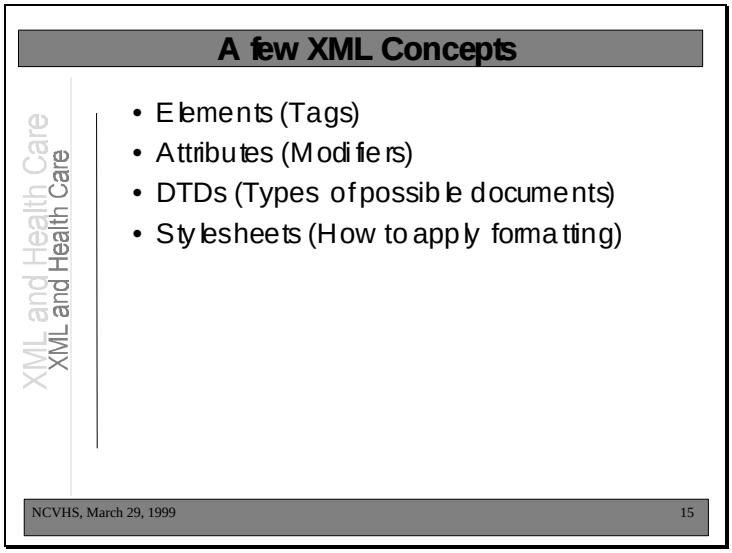
XML provides a mechanism for electronically encoding documents and types of documents. XML documents contain a list of elements or tags which appear in a document, a specification of the order and frequency of elements in the document, a verification that the document is XML and facility for specifying formatting through style sheets.

Slide 14

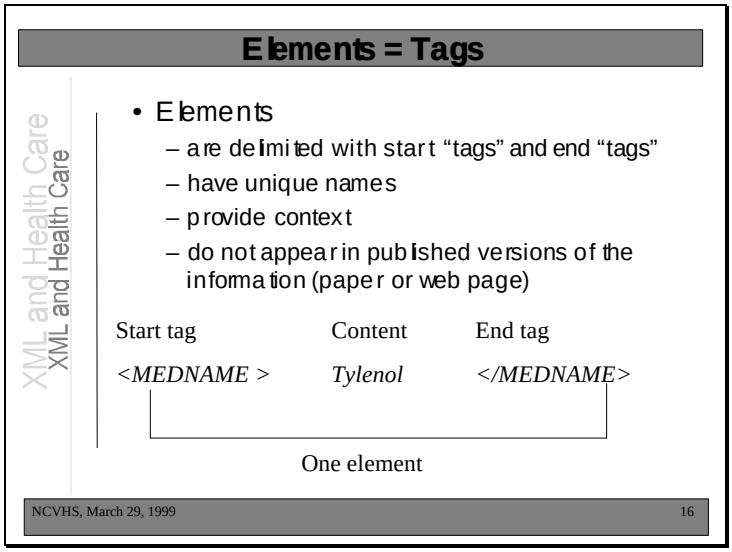


XML has the facility to present multiple characters in all foreign languages. XML will be used for international language exchange because it can handle different character sets such as Japanese and Korean. Microsoft is planning to include XML support in April, 1999. Previous DTP that support SGML currently support XML. SGML became a standard in 1986. SGML documents created 13 years ago can still be processed by any SGML-aware application today.

Slide 15

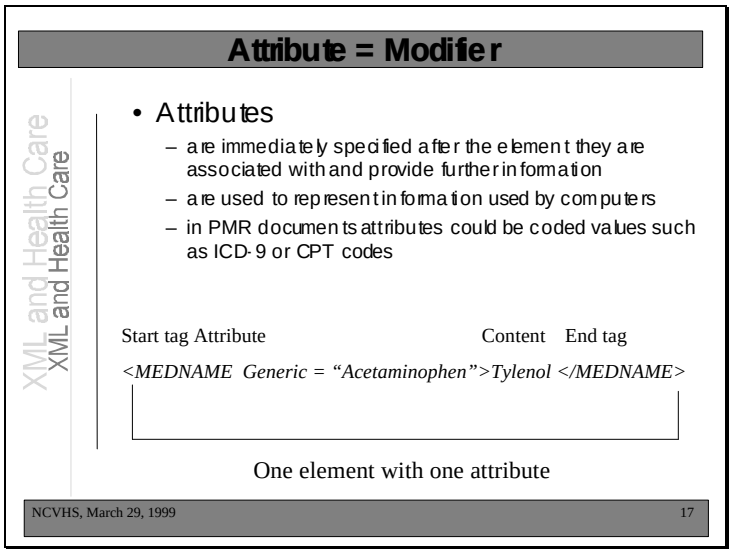


Slide 16



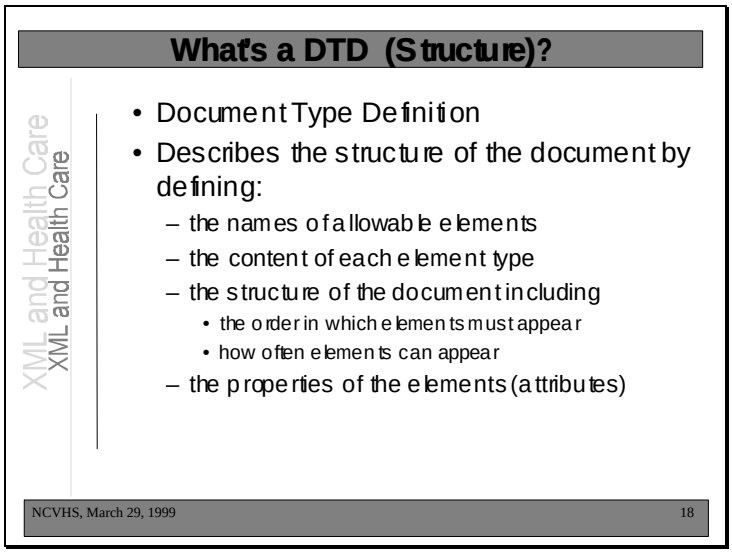
Elements are also known as "tags" and make up the mark-up of the information.

Slide 17



Attributes are further information about the content of the tag and are only readable by the computer.

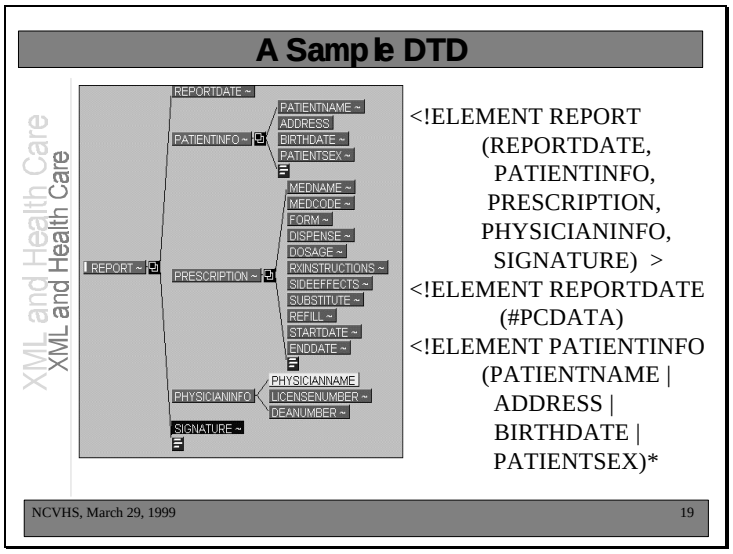
Slide 18



The definition of the type of document is the DTD. The names of the tags and the rules for using them are contained in the DTD or *Document Type Definition*. XML documents may or may not have an associated DTD.

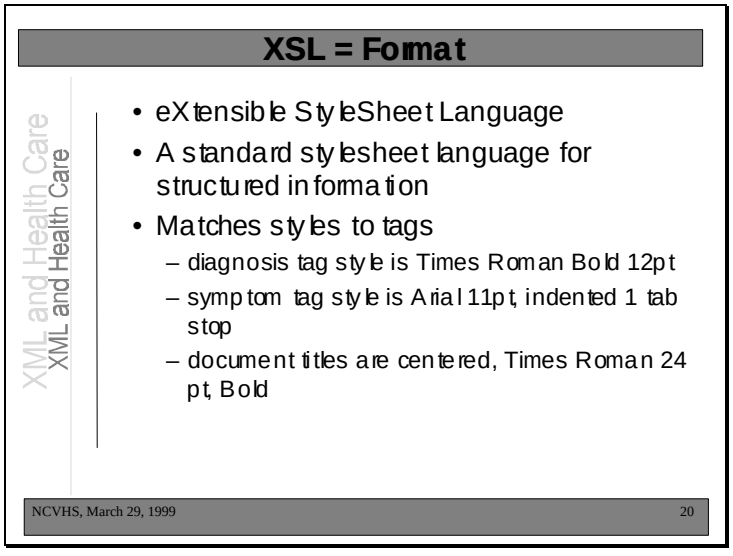
DTDs describe the document. A DTD is authored and can reflect as much or little as is desired. DTDs can describe any type of clinical document found in the Patient Medical Record such as a mammography report, encounter registration, a discharge summary. DTDs can describe other types of documents such as forms (HCFA 1500 form, CDC typhoid fever surveillance report, birth certificate) or other documents such as guidelines and protocols.

Slide 19



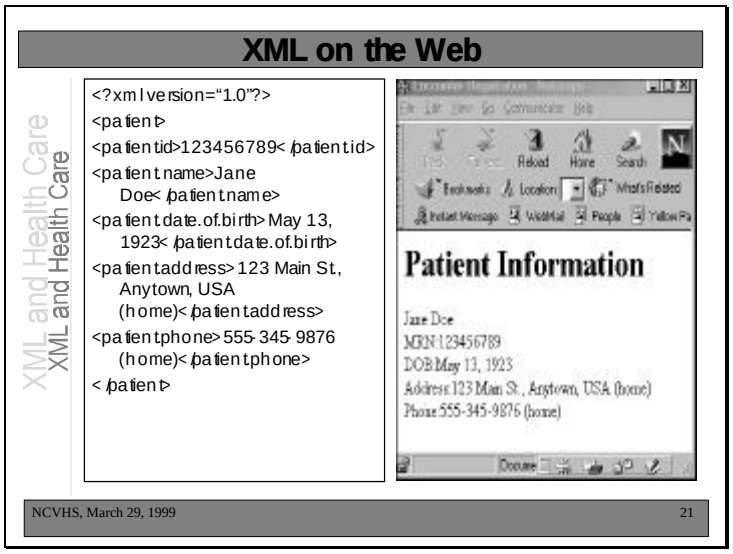
The DTD describes the structure of the document and defines the names of tags it contains. Additionally the DTD declares the order in which the tags occur and how often the tags can appear. A DTD for a prescription might contain structural elements for the medication prescribed, the dosage, the form, the quantity, etc. DTDs can describe documents in a clinical setting, such as a prescription. A prescription has a regular structure as represented in the DTD example: a date is followed by information about the patient, which followed by the actual prescription information, which is then followed by some information about the physician and a signature.

Slide 20



XSL is an acronym for eXtensible StyleSheet Language. It matches styles to tags in XML documents. For instance, the diagnosis tag style is Times Roman Bold 12pt, the symptom tag style is Arial 11pt, indented 1 tab stop and document titles are centered, Times Roman 24 pt, Bold.

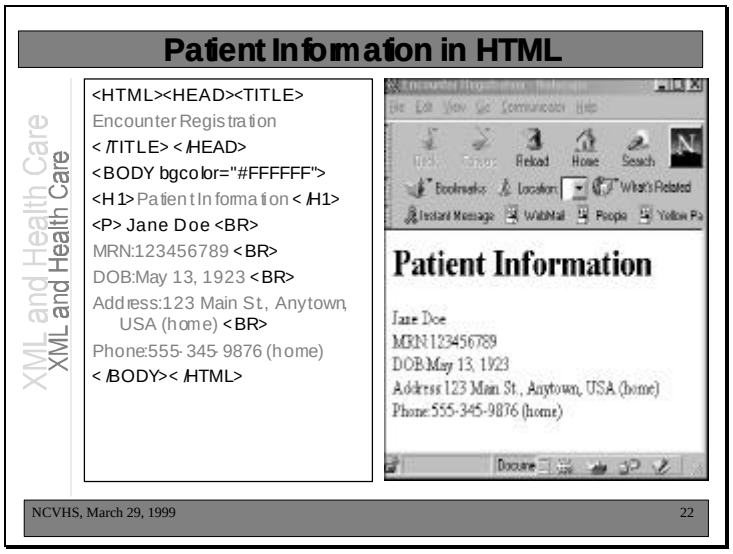
Slide 21



An XML document with a stylesheet for a web page, will generate an HTML document for web browsers.

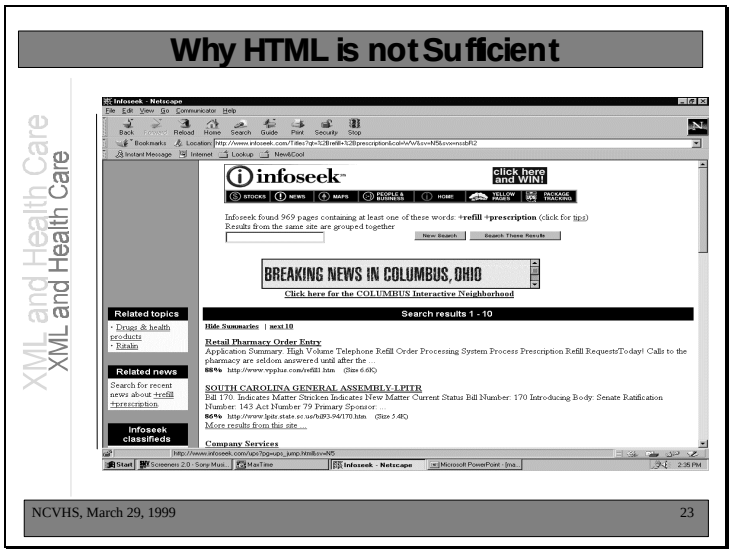
Internet Explorer 5.0 which was released about two weeks ago on March 18, 1999 has support for XML documents with XSL stylesheets.

Slide 22



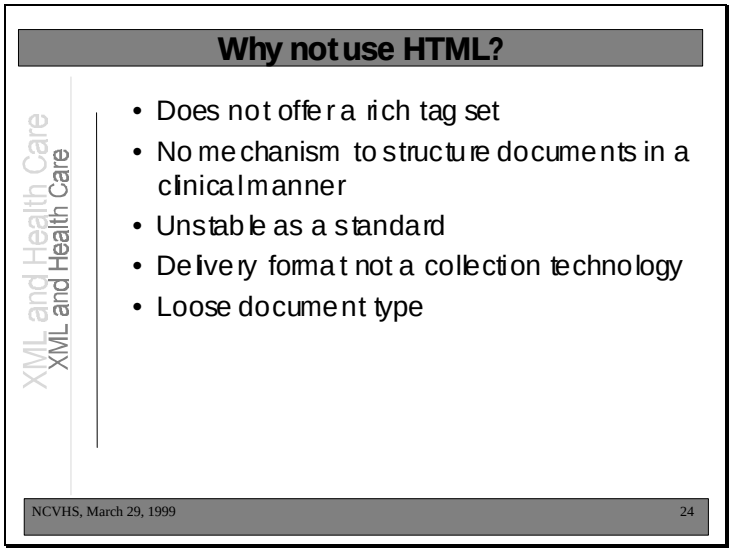
What is the relationship between HTML and XML? HTML is HyperText Markup Language. HTML is a mark-up language for describing web pages and contains tags that describe the format of information. HTML is SGML with a specific DTD. The HTML DTD has tags that are relevant to publishing information to a computer screen. These include headings <H1>, paragraph text <P> and a tags for linking.

Slide 23



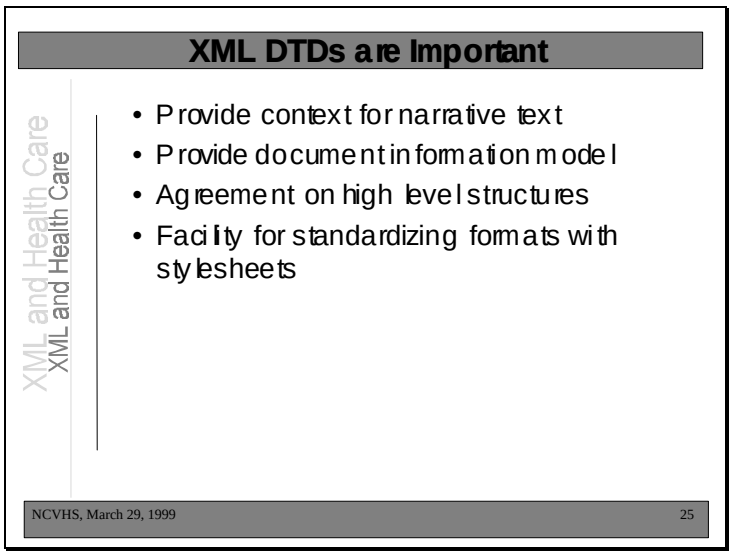
This difficulty in finding the context of information becomes especially apparent when surfing the web for information. Since HTML has a limited set of tags, it is difficult to locate specific information in the correct context. I performed the following search looking for "prescription" and "refill". As can be seen by the search results, my ability to find information about prescription and refills is not necessarily easy. In fact, I was presented with information about the South Carolina General Assembly which has nothing to do with the Patient Medical Record Information I had in mind.

Slide 24



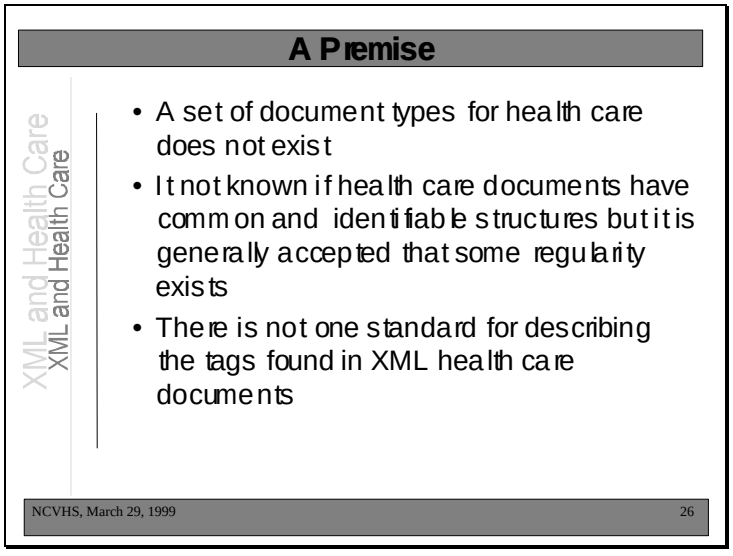
HTML is **not** a technology that meets the needs of information within the Patient Medical Record. The HTML standard or the HTML DTD is not a stable standard. HTML is currently up to HTML version 4.0 and many browser developers add non-standard extensions to HTML to compensate for some of its limitations.

Slide 25



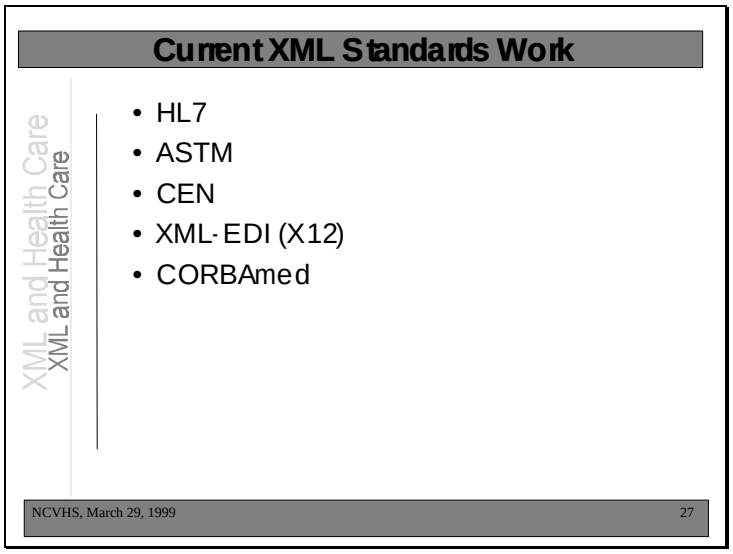
XML DTDs are important for Patient Medical record Information because they will provide context for narrative text, provide a document information model, allow for agreement on high level structures, and they will provide a facility for standardizing formats with stylesheets.

Slide 26



A standard set of DTDs for the Patient Medical record does not exist.

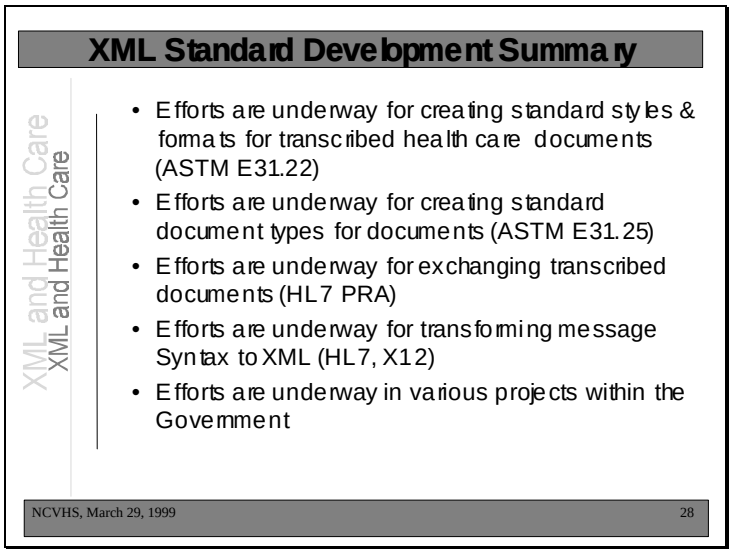
Slide 27



XML efforts have focused on:

- XML's relationship to object models: (HL7 V3, ASTM, CORBAmed, CEN),
- using XML as a syntax for health care messages and transaction: HL7 X12)
- document exchange (HL7 Patient Record Architecture PRA)
- financial transactions (XML-EDI, X12) and recently
- document types and formats (ASTM).

Slide 28



The efforts may be summarized as:

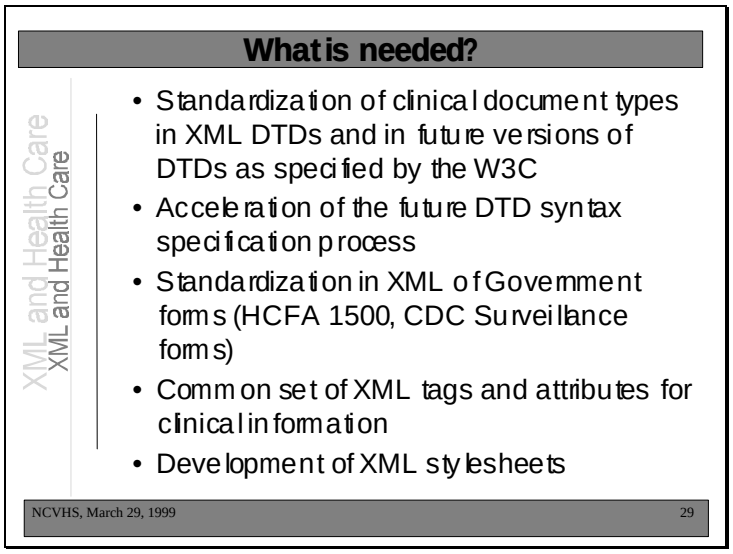
Frameworks for information exchange which includes HL7 and XML-EDI. Messages, transactions, and architectures to request and send healthcare information.

Services for healthcare information: CORBAmed, CEN and HL7. Interfaces to find, request, send, filter, and query healthcare information.

Research: CEN. Uses of XML in healthcare and best representations.

Paper-based Forms and Documents: regulatory forms for reporting healthcare information (HCFA, CDC) and documentation produced from the transcription process.

Slide 29



Efforts to create DTDs and sytlesheets for the documents in the Patient Medical Record are newly formed activities of ASTM and not much progress has been made in their short lifetimes. The new DTD syntax specified by the W3C will greatly impact the future of structuring XML document and the rules about constructing documents. Additionally, DTDs for common forms from the government should be available. There is a need for standardization among tag names. Different standards may create their own tag names and HL7 tag names might differ from ASTM. At the very least, a facility needs to be provided for mapping XML tags to different coding systems and to other tag names that mean the same thing.

Slide 30