Foundations of Data Curation—IS 513
Assignment 2

Adrian Bandolon

October 1, 2017

Write prose documentation for each element, attribute and attribute value.

- The root element is *progressNote,* and it must contain the elements *heading, history, careProvider* and *signature*. These elements can only occur once and must be stated in order. Each element within *progressNote* has type PCDATA.

- *heading* is a child element of *progressNote* and has the elements *date, time, name, patientNumber.* These elements can only occur once and must be stated in order. *heading* has a fixed attribute *"Visit Specifics"* of type CDATA. This attribute was added to serve as metadata to improve readability.
  - The *date* element is used as a child element in other parts of the document. The structure of the *date* element was generalized to minimize element redundancy. *date* must contain *day*, *month*, and *year* elements of type PCDATA. Each of these elements can only occur once. A required attribute for the *date* element is used to differentiate the instances where *date* can occur.
  - The *time* element is similar to the *date* element in that it is used in multiple parts of the document and it is structured similarly. *time* has the elements *hr*, *min*, *sec* and *timeOfDay* of type PCDATA. These elements can only appear once. *time* has a require attribute to differentiate the places where *time* can be used.
  - The *name* element is also used in multiple parts of the document. It has two child elements, *firstName* and *lastName*. These elements can only appear once. Since there are multiple instances of the *name* element, it has a required attribute to help differentiate the occurences.
  - *patientNumber* is a child of *heading* that can appear only once and it is of type PCDATA.

- The *history* element contains the elements *patientStatus, targetSymptoms, basicBehaviors, sideEffects, mentalStatus, specialCircumstances, vitalSigns, nursingInterventions, levelOfCare, linkToTheTreatmentPlan,* and *interventions.* These sub-elements can contain zero or more occurrences of type PCDATA. A fixed attribute *"Patient History and Intervention"* is included in this element for improved documentation.
  - *vitalSigns* is a child element of *history* that has sub-elements *sittingBP, sittingPulse, respRate, temp* and *weight*. These sub-elements can occur only once and are of type PCDATA. Fixed attributes are attached to these sub-elements to ensure the use of standard units of measure (UOM).

- *linkToTheTreatment* has sub-elements *problem, shortTermGoals,* and *date* of type PCDATA. These sub-elements can only appear once. The *date* element in this instance is the *target date* which can be differentiated using the *date* attributes described previously.

- The *careProvider* element has child elements *name* and *providerTitle* that can appear only once and is of type PCDATA. A fixed attribute of type CDATA is attached to *careProvider* to serve as metadata. *title1* and *title2* are entities that can make it easier to fill in and standardize the value used in *providerTitle*.

- *signature* element has sub-elements *method, name, providerTitle, date* and *time* that must appear once of type PCDATA. *signature* also has a fixed attribute *"Signature Specifics"* serving as metadata. All sub-elements except for method has been described earlier.

How did you decide to represent the data in the way that you did? Why did you choose the elements and attributes that you did?

- This DTD is designed with data independence in mind. I wanted this to be able to serve as a template for other nurse's notes. It needed to accommodate changes in the data without affecting the DTD structure. To this end, elements that can have multiple instances in the document are generalized (e.g. *date*, *time* and *name*). The instances of these elements are distinguished from each other using attributes. Attributes and comments serve as section dividers to improve readability.

What were the hardest decisions you had to make in this design process? How does you DTD design support data independence?

- The initial design was based on the document. With data independence in mind, as the project progressed new realizations were made that required changes to the initial design. The most challenging part of the project was the tedious work of going back to the DTD and modifying the elements and attributes that needed changing. I wish those design change decisions were realized initially, but that kind of foresight could only have come with experience, which I am still lacking.
- This design can serve as a template for other progress notes. The features in this design can accommodate changes to the data without affecting the DTD design itself. It is general enough to accommodate new information (e.g. *date, time, TargetSymptoms*, *mentalStatus*), yet has features that are specific to support standardization (e.g. *unit of measure* attributes)

How may your DTD design support the overarching goals of data curation?

- This design supports the overarching goals of data curation through the use of generalized elements that can be re-used (e.g. *data, time* and *name*) as well as features that requires specific parts to be standardized. This allows this design to be re-used as well as provide a structure that can dependably and efficiently support data analysis.