Structures and Standards in IOA

Dave Dubin

March 6, 2018

Standards as invisible solutions

- We benefit from standards without noticing them.

Standards as invisible solutions

- We benefit from standards without noticing them.
- Standards allow computers and computer programs to share information, even when the hardware or software has been designed by different individuals or companies.

Standards as invisible solutions

- We benefit from standards without noticing them.
- Standards allow computers and computer programs to share information, even when the hardware or software has been designed by different individuals or companies.
- Information processing is only one of many areas of our lives in which standards are important. For example, automobile parts and the voltage of household electrical current are standardized. Money is a standard medium of exchange.

What we mean by an information processing standard

By an "information standard" we may mean:

1. A method, protocol, or system for accomplishing a particular task, such as encoding information in a file or sharing it over a network.

What we mean by an information processing standard

By an "information standard" we may mean:

1. A method, protocol, or system for accomplishing a particular task, such as encoding information in a file or sharing it over a network.
2. Hardware or software that employs or executes that method or protocol, (for example, a word processing program).

What we mean by an information processing standard

By an "information standard" we may mean:

1. A method, protocol, or system for accomplishing a particular task, such as encoding information in a file or sharing it over a network.
2. Hardware or software that employs or executes that method or protocol, (for example, a word processing program).
3. A document that specifies the method or protocol in very detailed, precise technical language.

What we mean by an information processing standard

By an "information standard" we may mean:

1. A method, protocol, or system for accomplishing a particular task, such as encoding information in a file or sharing it over a network.
2. Hardware or software that employs or executes that method or protocol, (for example, a word processing program).
3. A document that specifies the method or protocol in very detailed, precise technical language.
4. An agreement among organizations or individuals that such a document represents.

Proposed definition

For our purposes, an adequate definition of "information processing standards" is that they are precisely documented agreements about methods or protocols for information processing, that are realized in the operation of computer hardware and software (Dubin, 2002).

Standards as solutions

- The impetus for the creation of any standard is to address a particular problem.

Standards as solutions

- The impetus for the creation of any standard is to address a particular problem.
- Consider the problem of representing the content of documents in computer files. The most basic problem is how to represent individual text characters as sequences of binary digits.

Standards as solutions

- The impetus for the creation of any standard is to address a particular problem.
- Consider the problem of representing the content of documents in computer files. The most basic problem is how to represent individual text characters as sequences of binary digits.
- The influential ASCII standard is very limited as a solution to this problem: 7 bits = 128 distinctions, 8 bits = 256.

Standards as solutions

- The impetus for the creation of any standard is to address a particular problem.
- Consider the problem of representing the content of documents in computer files. The most basic problem is how to represent individual text characters as sequences of binary digits.
- The influential ASCII standard is very limited as a solution to this problem: 7 bits = 128 distinctions, 8 bits = 256.
- The Unicode standard aims to include every major script in the world and every technical symbol in common use.

Standards as solutions

- The impetus for the creation of any standard is to address a particular problem.
- Consider the problem of representing the content of documents in computer files. The most basic problem is how to represent individual text characters as sequences of binary digits.
- The influential ASCII standard is very limited as a solution to this problem: 7 bits = 128 distinctions, 8 bits = 256.
- The Unicode standard aims to include every major script in the world and every technical symbol in common use.
- Unicode has proven to be a success story, despite its basis in a fundamentally flawed model.

ASCII Example

01000001	0x41	65	A	01010000	0x50	80	P
01000010	0x42	66	B	01010001	0x51	81	Q
01000011	0x43	67	C	01010010	0x52	82	R
01000100	0x44	68	D	01010011	0x53	83	S
01000101	0x45	69	E	01010100	0x54	84	T
01000110	0x46	70	F	01010101	0x55	85	U
01000111	0x47	71	G	01010110	0x56	86	V
01001000	0x48	72	H	01010111	0x57	87	W
01001001	0x49	73	I	01011000	0x58	88	X
01001010	0x4a	74	J	01011001	0x59	89	Y
01001011	0x4b	75	K	01011010	0x5a	90	Z
01001100	0x4c	76	L	01100001	0x61	97	a
01001101	0x4d	77	M	01100010	0x62	98	b
01001110	0x4e	78	N	01100011	0x63	99	c
01001111	0x4f	79	O	01100100	0x64	100	d

Latin 1 Example

01000001 0x41 65 A
01000100 0x44 68 D
01000101 0x45 69 E
01001010 0x4a 74 J
01100001 0x61 97 a
01100010 0x62 98 b
01100100 0x64 100 d
11000101 0xc5 197 Å
11100101 0xe5 229 å
11100110 0xe6 230 æ
11111000 0xf8 248 ø

UTF-8 Example

01000001 0x41 65 A
01000100 0x44 68 D
01000101 0x45 69 E
01001010 0x4a 74 J
01100001 0x61 97 a
01100010 0x62 98 b
01100100 0x64 100 d
1100001110000101 0xc385 50053 Å
1100001110100101 0xc3a5 50085 å
1100001110100110 0xc3a6 50086 æ
1100001110111000 0xc3b8 50104 ø

Standards documents are difficult to read

- Standards emerge from the work and consensus of many people, and they therefore represent solutions that are general enough to address a variety of problems.

Standards documents are difficult to read

- Standards emerge from the work and consensus of many people, and they therefore represent solutions that are general enough to address a variety of problems.
- The same solution (or nearly the same) may be published by different organizations under different names (e.g., ANSI X.3, ISO 646).

Standards as Agreements

There are various criteria for what counts as a standard:

1. A *de facto* standard is a solution that has become widely adopted, and is considered standard by virtue of its popularity.

Standards as Agreements

There are various criteria for what counts as a standard:

1. A *de facto* standard is a solution that has become widely adopted, and is considered standard by virtue of its popularity.
2. A *de jure* standard has been reviewed and formally approved by a Standards Developing Organization (SDO) such as ISO or one of its member organizations (e.g., ANSI in the United States).

Standards as Agreements

There are various criteria for what counts as a standard:

1. A *de facto* standard is a solution that has become widely adopted, and is considered standard by virtue of its popularity.
2. A *de jure* standard has been reviewed and formally approved by a Standards Developing Organization (SDO) such as ISO or one of its member organizations (e.g., ANSI in the United States).
3. There are public specifications similar to *de jure* standards, but authorized by industry consortia.

Criteria for standards

- Some technologies are deemed standards solely by virtue of their wide acceptance and popularity.

Criteria for standards

- Some technologies are deemed standards solely by virtue of their wide acceptance and popularity.
- Openness: standards are designed and documented with the aim of making every detail public.

Criteria for standards

- Some technologies are deemed standards solely by virtue of their wide acceptance and popularity.
- Openness: standards are designed and documented with the aim of making every detail public.
- Stability: there's often a tradeoff between the stability of strict adherence and the freedom to innovate.

Criteria for standards

- Some technologies are deemed standards solely by virtue of their wide acceptance and popularity.
- Openness: standards are designed and documented with the aim of making every detail public.
- Stability: there's often a tradeoff between the stability of strict adherence and the freedom to innovate.
- Consensus: *de facto* information standards represent a consensus among users that an existing application or protocol is worth adopting, while *de jure* standards are designed from the beginning to address as wide a range of needs as possible.

Markup

- Markup identifies the boundaries and properties of structures in a text.

Markup

- Markup identifies the boundaries and properties of structures in a text.
- Markup consists of symbols from the same repertoire as those used to encode the text.

Markup

- Markup identifies the boundaries and properties of structures in a text.
- Markup consists of symbols from the same repertoire as those used to encode the text.
- Even whitepace and punctuation are a kind of markup.

Markup

- Markup identifies the boundaries and properties of structures in a text.
- Markup consists of symbols from the same repertoire as those used to encode the text.
- Even whitepace and punctuation are a kind of markup.
- Presentational markup controls how document content is presented.

Markup

- Markup identifies the boundaries and properties of structures in a text.
- Markup consists of symbols from the same repertoire as those used to encode the text.
- Even whitepace and punctuation are a kind of markup.
- Presentational markup controls how document content is presented.
- Structural markup highlights properties that can be presented more than one way.

Markdown

```
**Assignments, Exercises & Grade Distribution:**

 - Assignment 1: Information Needs/Information Seeking
   Behavior. Due A1DUE (20%).
 - Assignment 2: Digital Collections Assessment.
   Due A2DUE (20%).
 - Assignment 3: Final Project. Due A3DUE (40%).
 - 10 Labs for Attendance and Completion/Class
   Participation (20%)

**Labs:**

 1. Library resources (LAB1DATE)
 2. ITD on computing resources (LAB2DATE)
 3. Research methods exercise (LAB3DATE)
 4. Pandoc encoding and transformation (LAB4DATE)
```

Generalized vs. specialized markup

- A particular markup language is fixed with respect to the meanings of the tags.

Generalized vs. specialized markup

- A particular markup language is fixed with respect to the meanings of the tags.
- A generalized markup framework allows you to define your own meanings.

SGML/XML Markup

```
<syllabus>
  <head>
    <course>
      <cname>Information Modeling</cname>
      <cnumber>LIS 561</cnumber>
      <dept>School of Information Sciences</dept>
      <term>Fall 2016</term>
    </course>
    <sections>
      <section>
         <secid>Section A</secid>
         <when>Monday 1:00 PM&ndash;3:50 PM</when>
         <where>Room 126, LIS Building</where>
      </section>
   </sections>
```

RDF as generalized markup

```
:SemesterFall2016 a :Semester ;
  event:time [ a tl:Interval ;
      tl:at "2016-08-22T00:00:00-5:00"^^xsd:dateTime ;
      tl:duration "P70D"^^xsd:duration] ;
  dcterms:subject :GenBackground ;
  rdfs:label "Fall semester"@en .

:Asgt3 a :Assignment ;
  rdfs:label "Assignment 3: Research Design Proposal"@en ;
  :moodleURL hwk:MOODLEASGT3 ;
  :sylSecId "Asgt3" ;
  :hasDeadline :A3M1, :A3M2, :A3M3, :A3D1, :A3D2, :A3D3 .

:A3M1 a :Deadline ;
  rdfs:label "Assignment 3 research question draft"@en ;
  :dueDuring :Week3 ;
  :dueDate "LAB3DATE" .
```

Further information

Dubin, D. (2002). Standards and information. In J. R. Schement (Ed.), *Encyclopedia of communication and information* (Vol. 3, pp. 965–967). New York: Macmillan.