

Developing International Electronic Data Exchange From Negotiations To Code

William S. Calvert, *FDA/CDER
Margaret K. Calvert, Service Research, Inc.

Abstract:

The development of electronic standards for international electronic data interchange (EDI) is described from a conceptual point of view. Practical considerations and implementation issues are discussed and clarified using SAS® code samples to read and write segments of standard data files. Examples will be drawn from standards in development, such as the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use expert working group on Data Elements for Transmission of Individual Case Safety Reports (ICH:E2b). Related implementation level standards such as those produced by the ICH:M2 expert working group and 'standards of convenience' such as Standard Generalized Markup Language (SGML), will also be discussed.

Introduction:

It has been observed and recorded that electronic and telecommunications technology is bring things closer together. As I write this we are observing the surface of a distant planet using a remotely controlled 6 wheel vehicle. Almost any commercial operation in developed countries has or soon will have at least casual connection to the Internet and global communication and information access.

In this global environment of communication, reliable and secure point to point exchange of data is becoming more common place. The pharmaceutical industry, for example consists of hundreds of firms many of which are large multi-national concerns. The way they do business is changing rapidly. Not only are internal communications changing, but the way they interact with regulators and other firms is changing as well. In a year or two, even those firms that are not currently using EDI will be starting to use it.

We all use standards. Standards work much like a common language. If we both use the same words and the words have the same meaning, we can exchange thoughts, ideas and information. We can work together effectively. The main advantage of standardization in the commercial sense is the

economies of scale. It is cheaper to build software and systems for EDI if everyone is using the same basic set of tools to send the same general message, using the same kinds of equipment. This single advantage has multiple ramifications: labor to keep the system going is easier to obtain and less expensive, communications and exchange of information is faster and easier to setup in a standardized environment, and consequently the overall responsiveness of the entire industry is enhanced. Responsiveness is notably important in the realm of post market drug safety monitoring, where the public health may be at risk.

Even though standards are recognized as a 'good' thing, they are not always widely used. Standards can be constraining. Using the language analogy, esoteric scientists often find it difficult to explain their work to the layman. Special words and concepts must be defined. There may be short-term commercial advantage in using advanced technology that others have not yet adopted. There is also the inertia from the way things have been done in the past. In the international arena, regional legacies can be very strong. Nationalism and the sense that my way is better, may impede progress towards a universally usable standard. There is also the need to design a standard for the 'lowest common denominator'. For example, not everyone has the latest computing equipment, gigs of disk, or

computing speed. Standards must work for a range of situations.

Necessarily, standards cannot be developed until there is a commonly recognized need. This requires a unity of purpose among a large number of people, firms, or other entities. The economic value of standards must be recognized and a process established to allow all parties to have appropriate input into the development effort. A standard that is not adopted and widely used is of limited value.

To assure effective implementation of a standard the process of development should lead to enthusiasm for the new standard and a clear understanding of the value of implementation. Sometimes it may be necessary to promote the adoption of standards through the use of political processes and laws or regulations. Ultimately the development effort must rely on a high level of technical expertise, though the need to apply a standard over a wide range of conditions may obscure some of the state of the art expertise that is involved.

Example from the pharmaceutical industry:

The pharmaceutical industry is a good example of the environment for developing standards. There have been efforts going on for years to develop and implement standards. Currently, many things are happening, and the future is bright for the realization of a standardized environment. The focus of our example will be on post market activities associated with electronic data interchange and the ICH initiatives in particular.

ICH and other standards development efforts in the pharmaceutical industry have been going on for many years. The multinational nature of the industry and the universal concern over safety and efficacy provide strong economic incentives to converge the requirements for registration and marketing of drugs and medicinal products. ICH is the process with the purpose of converging the requirements for everything from animal studies, clinical trials design and statistical analysis, pharmacokinetics studies, through reporting of post market individual safety reports (ISR).

The ICH process is directed by a Steering Committee that has six representatives, one for the regulatory authority and one for industry in each of the three regions; Japan, the United States, and Europe. The process consists of 5 steps. In Step 1, an industry representative is selected as rapporteur and a rough draft of the proposed standard is developed by a panel of experts representing each of the 6 entities. This draft is circulated outside the expert working group and revised and improved until all six representatives can sign that they agree to the draft achieving Step 2. The Step 2 document is published for comment and to meet legal requirements in some countries. In Step 3 the rapporteur responsibility is passed to a person representing a regulatory authority and comments on the Step 2 document are incorporated to produce the version of the standard that all 6 representatives can agree to and the three regulatory representatives can sign to achieve Step 4. On presentation to and acceptance by the steering committee, the proposed Step 4 document becomes an ICH standard. Step 5 is achieved in each jurisdiction when the local regulatory authority adopts and implements the standard.

Note that the ICH process can be laborious and may take several years to culminate in an implementation. Implementation and adoption are dependent on local political processes. At each step, 'established' positions may need to be renegotiated and the joint understanding clarified. Differences in the way things have been done often cause misunderstandings.

E2b is an example of an ICH standard that will have a significant impact on the way the pharmaceutical industry reports adverse drug events. It is expected to achieve Step 4 at the Fourth ICH conference in Brussels, July 1997. This standard defines the content or fields that must be available for reporting individual safety reports of adverse drug/medicine events.

The E2b message as defined in the most recent draft, contains two major sections, the administrative and identification information group of fields (Section 2A) and the information on the case (Section 2B). In the 'A' section there are three parts, identification of the case safety report, the primary source of the

information, and information on the sender and receiver of the case safety report. The 'B' section has 5 parts, patient characteristics, description of the reaction/event, results of tests and procedures relative to the investigation of the patient, description of the drugs (biologicals, etc.) pertinent to the case, and provision for narrative case summary information and additional notes about the case.

Each part contains numerous fields designed to allow the representation of data at whatever level of detail is available. For example, AGE can be captured as the birth date and event date, the number of days, months, or years; or as an age category (eg. child, adolescent, adult, elderly). The data structure that is implied by this standard provides for as many types of cases as can be conceived beforehand, so necessarily not all fields can be filled for any one case. Going back to the AGE example, only the single best available representation of AGE should be supplied.

To be useful, other standards must be borrowed or developed to handle the details of data representation and transmission. Some of these kinds of problems are handled by other related ICH standards. The most important ones for E2b are those standards produced by ICH:M1 and ICH:M2. The M1 expert working group has developed an international medical terminology which has one preferred term for each medical concept that they could identify. The terminology is hierarchical in structure and ranges from 26 system organ classes in the smaller categories, to tens of thousands of lower level terms that encompass most of the ways each medical concept may be represented in different regions.

Once the message definition is standardized, a range of technical issues must be addressed; the ICH:M2 expert working group has taken on most of these issues. First, there are several data models that can be used to represent the message. Most of the difficulty with the data model relates to optimizing the relational database design to support the very broad range of data that can be represented in an E2b message. The issues revolve around the most logical placement of fields, particularly repeating fields.

At a more detailed level, the coding and representation of individual data must be standardized. ISO standards have been used where possible, such as for countries and for date formats. The international medical terminology developed by ICH:M1 has already been mentioned. It will be used for the adverse reaction code and the code for drug indication. There are other areas, perhaps less important where improvements in the standard are needed. For example, dosage forms and routes of administration have been suggested but have not been as thoroughly developed as the medical terminology. In these two instances, the regional legacies and differences in usage interact making the technical issues more difficult.

An additional dimension of implementation is in the changing of the regulatory environment to accommodate the new standards and the technology it enables. In some cases implementation requires a new way of thinking that was not required or possible in the paper world. An example of this is in the electronic signature rule published by the FDA to take effect in August 1997. Other regulations may need to be changed to allow the official report of an event to be sent in electronically instead of on a specific form printed on a piece of paper. Implicit in these changes are the process changes. The same things need to be done to process reports, but they are done differently for paper than for electronic transmissions. When there is a public trust or responsibility for the health and safety of the public, changes are not made lightly or quickly.

The most difficult part of implementation of standards is the very change it entails. People are generally resistant to change. Ultimately the motivator for change is economic. Apparent realizable economic advantage must be associated with the change. It takes resources to adopt new standards. Note also that if the change will promote stability (ie. a long period of no change) then the change will be viewed more favorably. People want a standard that will last.

Now that we have considered the message and the detailed representation of the information it contains, we have to look the other way and consider the outer layers and electronic 'envelope' in which the

message will be transmitted. Since the envelope is generally transparent to the data user, it is mentioned just briefly. TEMPLAR® is a commercial software product that was selected as the initial working 'envelope' standard since it offered encryption and basic functionality that was available across all regions.

As is often the case, a standard must be geared to be usable by the lowest common denominator. This is especially true of the outer layers. The message must be standardized so that it can be passed universally between platforms. Since the advent of the personal computer, the American Standard for Computer Information Interchange (ASCII) has become the most pervasive cross platform standard. This was considered along with several other representational standards such as EBCDIC, HL7, UNEDIFACT, etc. None really provided the full functionality required, though UNEDIFACT may be used in the future when the E2b message can be defined and accepted by the standards body (in 12-18 months).

To speed the process, the Japanese representatives on the expert working group creatively proposed the use of SGML and did the initial work to represent the data in this form. The quick description of SGML is that it is a 'tagged' ASCII file. However, a little history will promote understanding of the scope of SGML. The Standard Generalized Markup Language (SGML) has been around for a long time. After some years of use, at IBM for example, it became an ISO standard in 1986. It is also interesting to note that it is the parent of HTML, one of the technological achievements that has made the Internet and the World Wide Web functionally useful.

SGML was adopted by the ICH:M2 expert working group as a standard of convenience. It was there and easily adapted to the requirements at hand. The ICH:M2 SGML implementation of ICH:E2b consists of a standard Data Type Descriptor (DTD) file (see Display 1 for an example of a DTD file) and compliant SGML data files (see Display 2 for an example of a SGML file). The DTD is structured according to the SGML language specifications and can be used by an SGML translator to read or write compliant files regardless of the data content. That

is, the data from a database or other structure can be put into standard format, moved via various electronic media to another desired location and translated into another data structure while maintaining the content of the data.

In version 7, SAS may be able to read E2b/SGML data files using the DTD. Currently, however, the DTD must be used as a guide to write the translation software in the SAS data step. Code examples can be provided once the Step 4 DTD for E2b is available. Currently it has been delayed and is expected by September 1997.

As a preview we have selected some programming solutions that will need to be implemented in SAS code to read or write E2b SGML data files. First, the tags and general file structure needs to be checked to be sure it is valid. Second there are large narrative and descriptive fields ranging from 1,000 to 10,000 bytes that will need to be read and stored. These fields will be of variable length. Third, there are repeating sets of fields, some of which are nested within other sets of repeating fields. Finally, there are data items that are represented by two fields. For example AGE has a unit field (ie. days, months, or years) and a value field (26, or some other number), or the dates have a format field (ie. CCYYMMDD) and a value field (19970708). Specific coded solutions will be triggered by tag recognition of the translation program.

Conclusion:

It is important to note that the historic differences in the way things are done can be the most difficult hurdle to overcome in the establishment of international standards. The first solutions to a problem often are focused on the specifics of the local situation. When the solution is put into practice and it works, there is an intense pride and sense of accomplishment. These feelings need to be subjugated to some extent to solve the problem again for the broader situation. Sometimes this is achieved through the very high levels of intellectual development and the global understanding of the problems as evidenced by the participants in the process. More often the process of developing standards is driven by clear and obvious economic

incentives such as the restriction on resources and an increasing need for efficiency. These conditions promote the need for working together, which requires standards.

The international flavor of commerce has been the source of the economic incentives. In the pharmaceutical industry, selling products in multiple countries can be an onerous undertaking. Even after drugs are approved for use, the post market safety reporting requirements are different from country to country. Sometimes the same information may need to be prepared in 20-30 ways and submitted at different times to meet all applicable legal requirements.

As a multinational firm attempts to comply with multiple legal requirements with some degree of reliability and effectiveness, a functional convergence occurs. The grouping of similar activities to facilitate compliance with legal requirements gives an insight to the industry that begs for standards to promote reliability, cost-effectiveness, and timeliness. Ultimately the industry develops a pervasive sense that the cost of adopting standards is a wise investment. Clearly this is happening in the pharmaceutical industry today.

The ICH:E2b standard as implemented by ICH:M2 for adverse drug event reporting has been presented as an example to show some of the general principles in a specific situation. Implementation details of this standard have been described briefly. More details on the implementation of electronic submission of adverse drug events by pharmaceutical manufacturers to the FDA can be obtained by calling Paul Reinstein at 301-827-3223 or on the FDA web page.

☐ This paper represents the opinions of the authors, no statements are to be construed as positions or policy of any government agency.

SAS is a registered trademark of SAS Institute Incorporated, Cary, NC.

Display 1: A Sample DTD (Data Type Descriptor) file.

```
<!DOCTYPE E2B [
<!-- PUBLIC "-//ICHM2//DTD E2B//EN"
"e2b_v011.dtd" -->
```

```
<!--
```

The DTDs issued by the ICH M2 group is public domain in nature.

No one can claim copyright on these DTDs.

No commercial distribution is allowed.

The ICH is not responsible for any damage or financial loss

as using these DTDs.

This version is tentative and changes are expected.

Changes will be made through ICH M2 process.

```
-->
```

```
<!ENTITY % ope.rec "ope ( app | rep | del ) app
">
```

```
<!ENTITY % act.rec "act ( rep | del ) rep">
```

```
<!ELEMENT e2b -- ( safetyreport* ) >
```

```
<!ELEMENT safetyreport - O ( safetyreportid
) >
```

```
<!ELEMENT safetyreportid - O (
#PCDATA,safetyreportsq* ) >
```

```
<!ATTLIST safetyreportid %ope.rec; >
```

```
<!ELEMENT safetyreportsq - O ( #PCDATA,
(reportcountry? &
occurcountry? &
transmissiondateformat? &
transmissiondate? &
reporttype? &
serious? &
seriousnessdeath? &
seriousnesslifethreatning? &
seriousnesshospitalization? &
seriousnessdisabling? &
seriousnesscongenitalanomaly? &
seriousnessother? &
receivedateformat? &
receivedate? &
receiptdateformat? &
receiptdate? &
additionaldocument? &
documentlist? &
```

```

fulfillexpeditecriteria? &
authoritynumb? &
companynumb? &
othernumb? &
duplicate? &
casenullification? &
nullificationreason? &
(reportduplicate? &
linkedreport? &
primarysource? &
sender? &
receiver? &
patientinformation? ))) >
<!ATTLIST safetyreportsq %ope.rec; >

```

```

<!ELEMENT reportduplicate - O
(reportduplicatesq*) >
<!ELEMENT reportduplicatesq - O (#PCDATA,
(duplicatesource? &
duplicatenumb? )) >
<!ATTLIST reportduplicatesq %ope.rec;
>

```

```

<!ELEMENT linkedreport - O
(linkedreportsq*) >
<!ELEMENT linkedreportsq - O (#PCDATA,
linkreportnumb?) >
<!ATTLIST linkedreportsq %ope.rec; >

```

```

<!ELEMENT primarysource - O
(primarysourcesq*) >
<!ELEMENT primarysourcesq - O (#PCDATA,
(reportertitle? &
reportergivename? &
reportermiddlename? &
reporterfamilyname? &
reporterorganization? &
reporterdepartment? &
reporterstreet? &
reportercity? &
reporterstate? &
reporterpostcode? &
reportercountry? &
qualification? &
medicallyconfirm? &
literaturereference? &
studyname? &
sponsorstudynumb? &
observestudytype? )) >
<!ATTLIST primarysourcesq %ope.rec; >

```

```

<!ELEMENT sender - O
(sendertype? &
senderorganization? &
senderdepartment? &
sendertitle? &
sendergivename? &
sendermiddlename? &
senderfamilyname? &
senderstreet? &
sendercity? &
senderstate? &
senderpostcode? &
sendercountry? &
sendertel? &
sendertelexextension? &
sendertelcountrycode? &
senderfax? &
senderfaxextension? &
senderfaxcountrycode? &
senderemailaddress? ) >
<!ATTLIST sender %ope.rec; >

```

```

<!ELEMENT test - O (testsq*) >
<!ELEMENT testsq - O (#PCDATA,
(testdateformat? &
testdate? &
testname? &
testresult? &
testunit? &
lowtestrange? &
hightestrange? &
moreinformation? )) >
<!ATTLIST testsq %ope.rec; >

```

```

<!ELEMENT drug - O (drugsq*) >
<!ELEMENT drugsq - O (#PCDATA,
(drugcharacterization? &
medicinalproduct? &
obtaindrugcountry? &
drugbatchnumb? &
drugauthorizationnumb? &
drugauthorizationcountry? &
drugauthorizationname? &
drugstructuredosagenumb? &
drugstructuredosageunit? &
drugseparatedosagenumb? &
drugintervallosageunitnumb? &
drugintervallosagedefinition? &

```

```

drugcumulativedosagenumb?      &
drugcumulativedosageunit?      &
drugdosagetext?                 &
drugdosageform?                 &
drugadministrationroute?        &
drugparadministration?          &
drugindication?                 &
drugstartdateformat?            &
drugstartdate?                  &
drugstartperiod?                &
drugstartperiodunit?            &
druglastperiod?                 &
druglastperiodunit?             &
drugenddateformat?              &
drugenddate?                    &
drugtreatmentduration?          &
drugtreatmentunitduration?      &
actiondrug?                     &
drugrecurreadministration?      &
drugrecuration?                 &
drugadditional?                 &
(activessubstance?              &
 drugreactionrelatedness? ))) >
<!ATTLIST drugsq %ope.rec; >
.
.
.
<!ELEMENT narrativeincludeclinical - O
(#PCDATA, old?) >
<!ATTLIST narrativeincludeclinical %act.rec;
>

<!ELEMENT reportercomment - O
(#PCDATA, old?) >
<!ATTLIST reportercomment %act.rec; >

<!ELEMENT senderdiagnosis - O
(#PCDATA, old?) >
<!ATTLIST senderdiagnosis %act.rec; >

<!ELEMENT sendercomment - O
(#PCDATA, old?) >
<!ATTLIST sendercomment %act.rec; >
]>

```

Display 2: A sample SGML (Standard Generalized Markup Language) file.

```

<E2B>
<SAFETYREPORT>
<SAFETYREPORTID OPE=APP>001
<SAFETYREPORTSQ>001

<REPORTCOUNTRY>US
<OCCURCOUNTRY>US
<TRANSMISSIONDATEFORMAT>102
<TRANSMISSIONDATE>19970101
<REPORTTYPE>1
<SERIOUS>1
<SERIOUSNESSDEATH>1
<SERIOUSNESSLIFETHREATNING>1
<SERIOUSNESSHOSPITALIZATION>1
<SERIOUSNESSDISABLING>2
<SERIOUSNESSCONGINITALANOMALI>2
<SERIOUSNESSOTHER>2
.
.
.
<TEST>
<TESTSQ>001
<TESTDATEFORMAT>102
<TESTDATE>19951028
<TESTNAME>EOSINOPHILS
<TESTRESULT>.1
<TESTUNIT>%
<LOWTESTRANGE>0.01
<HIGHTESTRANGE>1
<MOREINFORMATION>2
</TESTSQ>
<TESTSQ>002
<TESTDATEFORMAT>102
<TESTDATE>19951031
<TESTNAME>EOSINOPHILS
<TESTRESULT>0.9
<TESTUNIT>%
<LOWTESTRANGE>0.1
<HIGHTESTRANGE>3
<MOREINFORMATION>2
</TESTSQ>
</TEST>

<DRUG>
<DRUGSQ>001
<DRUGCHARACTERIZATION>1
<MEDICINALPRODUCT>ASPRIN

```

<OBTAINDRUGCOUNTRY> US
 <DRUGBATCHNUMB> A1234
 <DRUGAUTHORIZATIONNUMB> B1234
 <DRUGAUTHORIZATIONCOUNTRY> US
 <DRUGAUTHORIZATIONNAME> BONES
 PHARMACEUTICALS
 <DRUGSTRUCTUREDOSAGENUMB> 6
 <DRUGSTRUCTUREDOSAGEUNIT> MG
 <DRUGSEPARATEDOSAGENUMB> 3
 <DRUGINTERVALDOSAGEUNITNUMB> 3
 <DRUGINTERVALDOSAGEDEFINITION> DAY
 <DRUGCUMULATIVEDOSAGENUMB> 6
 <DRUGCUMULATIVEDOSAGEUNIT> MG
 <DRUGDOSAGETEXT> 6MG OF ASPRIN 3
 TIMES A DAY FOR 2 DAYS
 <DRUGDOSAGEFORM> PILL
 <DRUGADMINISTRATIONROUTE> ORA
 <DRUGPARADMINISTRATION> ORA
 <DRUGINDICATION> HEART PROBLEMS
 <DRUGSTARTDATEFORMAT> 610
 <DRUGSTARTDATE> 199601
 <DRUGSTARTPERIOD> 2
 <DRUGSTARTPERIODUNIT> 803
 <DRUGLASTPERIOD> 0
 <DRUGLASTPERIODUNIT> 807
 <DRUGENDDATEFORMAT> 610
 <DRUGENDDATE> 199604
 <DRUGTREATMENTDURATION> 4
 <DRUGTREATMENTUNITDURATION> 802
 <ACTIONDRUG> 1
 <DRUGRECURREADMINISTRATION> 1
 <DRUGRECURACTION> TENSION
 HEADACHES
 <DRUGADDITIONAL> NONE
 <ACTIVESUBSTANCE>
 <ACTIVESUBSTANCESQ> 01
 <ACTIVESUBSTANCENAME> CHEMICAL
 COMPOUNDS
 <DRUGREACTIONRELATEDNESS>
 <DRUGREACTIONRELATEDNESSSQ> 01
 <DRUGREACTIONASSES> TENSION
 HEADACHES
 <DRUGASSESSMENTSOURCE> CLINICAL
 TESTS
 <DRUGASSESSMENTMETHOD> BLOOD
 TESTS
 <DRUGRESULT> POSIITVE
 </DRUGSQ>
 </DRUG>

<SUMMARY>
 <NARRATIVEINCLUDECLINICAL> FEMALE
 PATIENT CAME IN COMPLAINING OF PAIN
 AND WAS GIVEN A DRUG WITH WHICH SHE
 HAD A REATION
 <REPORTERCOMMENT> NONE
 <SENDERDIAGNOSIS> NONE
 <SENDERCOMMENT> NONE
 </SUMMARY>
 </SAFETYREPORTID>
 </SAFETYREPORT>
 </E2B>

Contact the authors by post or phone at:

Service Research Inc.
 401 Hull Place
 Rockville, MD 20852
 301-738-1024

or by email at:

CALVERTW@CDER.FDA.GOV