

DTD/Schema Project Definition Questionnaire

#	Question	Your Answer
<i>Organization name:</i> UBL Group/UBL TC		
<i>Interviewees:</i> UBL Planning Subcommittee		
<i>Date of interview:</i> 11 October 2001 (Q1-Q10), 18 October (Q11-Q25), 25 October (rest)		
General Project Goals, Scope, and Constraints		
1	What problems does your organization experience with its current environment that you want to address in this project? What are their priorities?	Crisis in interoperability and promulgation of electronic B2B commerce internationally. Existing technologies don't take advantage of modern programming techniques or XML, and are too expensive, complicated, and hard to set up. They're well set up for pairwise agreements, but not loosely coupled arbitrary agreements. See the charter for the group.
2	What information ("document types") is in the scope of this project? What is not in the scope?	See the October 11 minutes.
3	How much information (in "pages" or other measurement) is in the scope?	Less than 20% of international commerce is done with EDI. We'd like to satisfy the needs of 80%!
4	What sort of information is in the scope: text, tables, graphics, equations, fielded data, video, hyperlinks, etc.? In what proportions? Which parts of the information represent the most valuable investment?	Structured text (fielded data) is the bulk of it. Graphics and other non-text is usually exchanged by other means. Product catalogs tend to have images or video involved, and might have some "free narrative text", but most don't have it. Binary data is usually problematic to exchange in current systems. We want to distinguish between "content as product" (such as syndicated content, e.g. novels or videos) and "content as process". We're concerned with the latter. As for hyperlinks, "content as product" might have them, but the stuff we're mostly concerned with will just use links as a mechanism.
5	In how many languages is the information written? Does any of the information need to contain text in multiple languages?	Again, product catalogs are more likely to contain >1 natural languages than other kinds of messages are. On average, any one message is likely to contain only one natural language. Mavis: For the EU tendering process (bid to purchase on behalf of the EU), some things have to be in both French and in English. But this apparently is handled currently as two instances.
6	Under what constraints must the project work: deadlines, software tools that have already been chosen, requirements for interchange file formats, availability of key personnel, etc.?	The TC's schedule and available personnel impose constraints. We want to finish Phase 1 within about 12 months. We agreed to start with xCBL. The UBL Schema subcommittee recommended to develop (or at least deliver!) the schemas in W3C XML Schema form as a minimum; this hasn't been agreed on yet, but we'll posit the existence of this constraint.
Existing Processes and Tools		

7	How and with what tools/markup is your information being created now?	Various and sundry. Mom/pop shops use Notepad, even for heavily fielded data, and they sell one thing. Large corporations use ERP systems interfaces. In the middle, someone enters the data into form fields on a web page. Web services might generate messages entirely automatically.
8	How and with what tools is the information being managed now?	Various and sundry again. Mom/pop shops use files on disk or paper files in folders or wirebound ledgers. Courier motorbikes are used for conveying messages.
9	In what forms is the information being delivered or used now, and how are these forms created?	Various and sundry again. The lowest common denominator is really low.
10	Does existing information need to be converted permanently to the new XML form? If so, on what schedule, in what proportion to newly created information, and with what tools?	We don't know who has information that needs to be converted, but it's not our problem. We are dealing with an interchange format, so the notion of "permanent" conversion is out of scope. We don't ourselves own any data that needs to be converted, and companies (e.g., ERP systems) can compete to come up with good solutions to this.
Information Creation, Management, and Workflow		
11	Are you planning to reengineer the information content and structure at the same time as migrating to XML: making it more modular, making it more hyperlinked, applying a new writing methodology, etc.?	Since UBL is an interchange format, the question could be seen as: Are you willing to change the import/extract "interfaces" to everyone's applications? No, we're not asking anyone to change their systems underneath; we're just adding an interchange format. For anyone who uses UBL as their native format, great! But this isn't necessarily the intent.
12	What information creation and management tools are likely to be used in the new environment?	Various and sundry, again. With the advent of UBL, the only kinds of upgrades we expect to see in the creation and management tools will include lots more XML-based programming, including XSLT transformation. But all the existing tools will continue to be used too.
13	Who (or what) will be responsible for applying XML markup to the information? If humans are involved, what is their level of tools and markup knowledge and responsibility compared to their subject matter knowledge and responsibility?	Various and sundry, again.
14	Do you receive any XML source files from external sources?	Yes, companies might receive XML files in any of the existing dialects: xCBL, cXML, OAGIS, VCML, etc., etc. UBL should be convertible from and to these formats.
15	Will you need to convert non-XML files to XML form on a routine basis?	Yes, traditional EDI, Spec 2000, Excel spreadsheets, various industry-specific formats, tons of proprietary formats. UBL should be convertible from and to these formats.
16	How much influence and control can you exert over the quality of the XML markup?	None, for the stuff that comes in from outside. Lots, for the stuff that you produce.
17	If human authors are involved, what is the authors' current level of understanding/acceptance of XML and the new environment?	N/A

18	What are the minimum revisable units (MRUs)? Do these "chunks" also serve as reusable units? Retrievable units?	In ebXML, core components can be base-level or aggregate. The aggregate ones and business entities may be MRUs. For our purposes, the analog to content management MRUs may be the reusable schema modules that are in our repository. A context driver descriptor would be applied to get the desired modules (which might be as big as a whole document type, or any level below).
19	Will data from a database contribute to the content?	Yes.
20	If the information contains hyperlinks, which links will be able to be generated? Which must be manually authored?	We've already discussed hyperlinks above. Catalogs might have them, but otherwise no.
21	How many human authors work on a single delivered document? How many delivered documents are assigned to a single author?	N/A.
22	How much of the information is newly created each time versus revised? How much time is allowed for a revision cycle?	Various and sundry. For example, a purchase order might be taken in and adapted into a purchase order response, which is a different document type with many similar components. Since SMEs are an important target audience, we want to be able to demonstrate that this scenario is possible.
23	Who reviews the information? On what cycle? With how much control and formality?	N/A
24	Do you have other comments on information creation and management requirements?	N/A
Information Processing, Delivery, and Access		
25	What processing do you intend to perform on the information: formatting, indexing for online navigation, transformation to other DTDs/schemas or other data formats, extraction/assembly, translation, content analysis, etc.? What are the output formats and their relative priorities? What tools are likely to be used?	Various and sundry! There will be a lot of emphasis on processing for database loading and extraction, but processing for display will have a role too.
26	Do you need to deliver XML files anywhere? If so, do they need to conform to an interchange DTD/schema over which you have no control? If so, what is it?	We <i>are</i> the hub interchange schema.
27	Do you need to generate Braille or other output optimized for the print-disabled?	No. This is dealt with at lower application levels. Even if we broaden the question to deal with the general notion of device independence, we still don't think we need to worry too much about it because we're not targeting this data for a specific device. That said, graphics should required alternate text, as is considered good practice for HTML.
28	How often are deliveries made for each type of output?	Various and sundry.
29	How can the information be searched or navigated in each delivered form: by page number, table of contents, index, full-text search, keyword search, context-based search, hyperlinks, cross-references, etc.?	We briefly discussed Topic Maps as a way to perhaps navigate among many documents, we concluded that this question is not really applicable because this is an interchange format.

30	Do you have other comments on information processing, delivery, or navigation requirements?	No.
Analysis Input		
31	Are there any relevant existing DTDs/schemas or data formats (proprietary or standard) that address any part of the information in this project's scope?	xCBL 3.0, the two major EDI formats (UN/EDIFACT and X12), IDOC (the SAP format), OAGI, the Joint Core Components work, and RosettaNet are the obvious ones. SimpleB (formerly SimpleEDI) could be useful. VCML and the German DIN specification are fairly faithful representations of EDIFACT and/or X12 and could be useful to examine in this light; we could use them as clarified forms of EDI against which we can make queries. (It's "standards input" rather than "usage input".) There are too many domain-specific efforts to mention here, but their message implementation guides and the EDIFACT and X12 community guidelines may be useful to examine. OBI, AIAG, GCI, Bolero, XBRL, IFX, and SWIFT all have a lot of good artifacts. We will rely on UBL members' knowledge to suggest other formats to examine. We expect that the subcommittees will fill out matrices of document type constructs, so that they can reveal both the coverage of semantics and the (arbitrary or important) differences in structure. (The Mapping subcommittee is coming up with material that will help the design subcommittees do their work.) This kind of analysis will inform decisions about how to build in extensibility. xCBL has already gone through this exercise, so maybe it's not as bad as it seems!
32	Does thorough documentation exist for the current markup language, templates, and/or information creation processes?	Yes, but all in different forms, as noted above. xCBL is documented with various guides and guidelines that should be examined.
33	In what form are sample documents and other analysis input available?	XML, Word, etc.
34	What other analysis input are you able to provide? (E.g., project plans, standards, style guides, bug reports, retrieval queries.)	xCBL has a bunch of artifacts from its design work, but it's messy enough that we should ask Commerce One to provide it on a one-by-one basis. Also, it was noted that CBL 1.0 design rules are worthwhile to look at.
Focus and Design Principles		
35	Are you planning to use XML validation for different purposes at various stages of production: conversion, creation, electronic review, intermediate transformation, final transformation, etc.? Are you prepared to perform any necessary XML-to-XML transformations?	In a sense, yes; you turn the information in document type A in stage n of the process into document type B in stage n+1 of the process. (Audit trails capture "the same" information as it moves through the process.) The core library approach enables the information to be identifiably "the same" as it moves through. xCBL 3.0 used the RosettaNet PIP approach: If you recognize one leg of the process, it uses one document type. A possible design principle: "In a choreography, each transmission is a document type." Individually identified document types may proliferate, but they're not as important as the components inside the documents. "Master types" make all fields available, but you might have different permutations of required/optional/forbidden field patterns in order to make different document types. (Date order sent, date

		order received, date delivery sent, date delivery received.) The nut of the principle is reuse of components in multiple document types. It's useful to have multiple document types because you want to validate as early as possible, and not leave it to downstream applications (all fields optional, loosey-goosey style).
36	How important is DTD/schema prescriptiveness?	It's good, to the extent possible. But it's always a balancing act because some communities might not use or want certain fields. An interchange schema sometimes has to be looser than any one community might want.
37	How important is making a controlled variant of another DTD/schema?	UBL doesn't have to be, e.g., a subset or controlled extension of xCBL.
38	How important is content-based markup versus structural markup?	Things like catalogs and accounting information tend to be "meta-schemas" rather than schemas; you need to generate your own schema for these things. For transactional information, content-based is the rule.
39	How important is presentation independence?	Very. However, we do expect people to view these documents after they (the documents, not the people!) have been transformed.
40	How important is making the "right" design decisions versus making fast design decisions?	The priority is on urgency (the charter that talks about an impending crisis).
41	How important is XML compliance (in the DTD/schema and/or in the instance)?	100%.
42	Do you have a requirement or desire to use a non-DTD schema language? If so, which one?	XML Schema has been recommended by a previous subcommittee.
43	What other DTD/schema characteristics are important to you? (E.g., markup naming, modularity, parameterization, architectural forms.)	See all the other notes about schema design rules and naming conventions. Also, extensibility design will be very important.