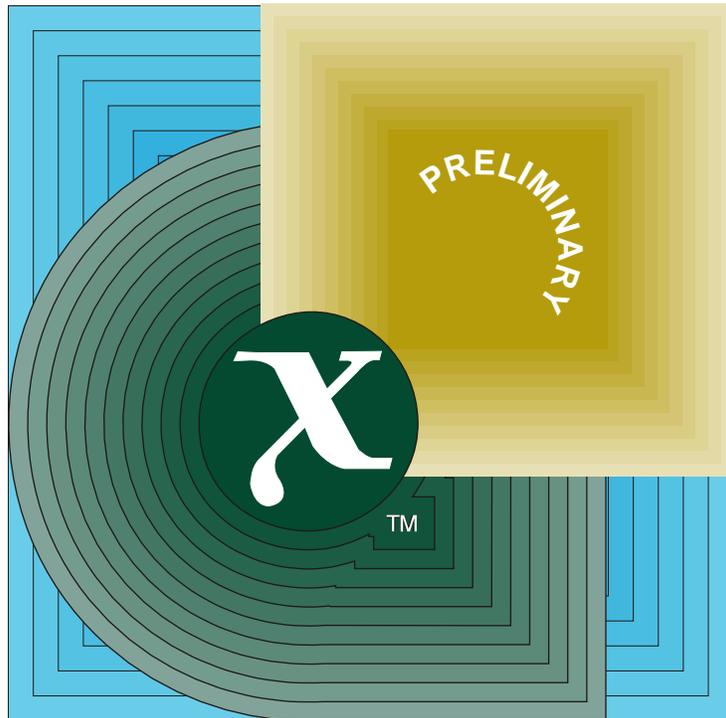


Preliminary Specification

Document Interchange Formats



THE *Open* GROUP

[This page intentionally left blank]

X/Open Preliminary Specification

Document Interchange Formats

X/Open Company Ltd.



© *September 1992, X/Open Company Limited*

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owners.

X/Open Preliminary Specification

Document Interchange Formats

ISBN: 1 872630 51 0

X/Open Document Number: P217

Set in Palatino by X/Open Company Ltd., U.K.

Published by X/Open Company Ltd., U.K.

Any comments relating to the material contained in this document may be submitted to X/Open at:

X/Open Company Limited
Apex Plaza
Forbury Road
Reading
Berkshire, RG1 1AX
United Kingdom

or by Electronic Mail to:

XoSpecs@xopen.co.uk

Contents

Chapter	1	Introduction	1
	1.1	Background.....	1
	1.2	Role of Reference Model.....	1
Chapter	2	ODA Interchange	3
	2.1	The ODA Standard.....	3
	2.2	Definitions.....	3
	2.3	Document Structure.....	4
	2.4	Compliance.....	4
	2.5	Limitations.....	5
Chapter	3	SGML Interchange	7
	3.1	The SGML Standard.....	7
	3.2	Definitions.....	7
	3.3	Document Structure.....	8
	3.4	Compliance.....	8
	3.5	Limitations.....	8
Chapter	4	Effects on Implementations	9
	4.1	Publication Requirements.....	9
	4.2	Non-requirements.....	9
	4.3	Testing Compliance.....	9
Appendix	A	Glossary	11

Preface

X/Open

X/Open is an independent, worldwide, open systems organisation supported by most of the world's largest information systems suppliers, user organisations and software companies. Its mission is to bring to users greater value from computing, through the practical implementation of open systems.

X/Open's strategy for achieving this goal is to combine existing and emerging standards into a comprehensive, integrated, high-value and usable system environment, called the Common Applications Environment (CAE). This environment covers the standards, above the hardware level, that are needed to support open systems. It provides for portability and interoperability of applications, and allows users to move between systems with a minimum of retraining.

The components of the Common Applications Environment are defined in X/Open CAE Specifications. These contain, among other things, an evolving portfolio of practical application programming interfaces (APIs), which significantly enhance portability of application programs at the source code level, and definitions of, and references to, protocols and protocol profiles, which significantly enhance the interoperability of applications.

The X/Open CAE Specifications are supported by an extensive set of conformance tests and a distinct X/Open trademark - the XPG brand - that is licensed by X/Open and may be carried only on products that comply with the X/Open CAE Specifications.

The XPG brand, when associated with a vendor's product, communicates clearly and unambiguously to a procurer that the software bearing the brand correctly implements the corresponding X/Open CAE Specifications. Users specifying XPG-conformance in their procurements are therefore certain that the branded products they buy conform to the CAE Specifications.

X/Open is primarily concerned with the selection and adoption of standards. The policy is to use formal approved *de jure* standards, where they exist, and to adopt widely supported *de facto* standards in other cases.

Where formal standards do not exist, it is X/Open policy to work closely with standards development organisations to assist in the creation of formal standards covering the needed functions, and to make its own work freely available to such organisations. Additionally, X/Open has a commitment to align its definitions with formal approved standards.

X/Open Specifications

There are two types of X/Open specification:

- *CAE Specifications*

CAE (Common Applications Environment) Specifications are the long-life specifications that form the basis for conformant and branded X/Open systems. They are intended to be used widely within the industry for product development and procurement purposes.

Developers who base their products on a current CAE Specification can be sure that either the current specification or an upwards-compatible version of it will be referenced by a future XPG brand (if not referenced already), and that a variety of compatible, XPG-branded systems capable of hosting their products will be available, either immediately or in the near future.

CAE Specifications are not published to coincide with the launch of a particular XPG brand, but are published as soon as they are developed. By providing access to its specifications in this way, X/Open makes it possible for products that conform to the CAE (and hence are eligible for a future XPG brand) to be developed as soon as practicable, enhancing the value of the XPG brand as a procurement aid to users.

- *Preliminary Specifications*

These are specifications, usually addressing an emerging area of technology, and consequently not yet supported by a base of conformant product implementations, that are released in a controlled manner for the purpose of validation through practical implementation or prototyping. A Preliminary Specification is not a “draft” specification. Indeed, it is as stable as X/Open can make it, and on publication has gone through the same rigorous X/Open development and review procedures as a CAE Specification.

Preliminary Specifications are analogous with the “trial-use” standards issued by formal standards organisations, and product development teams are intended to develop products on the basis of them. However, because of the nature of the technology that a Preliminary Specification is addressing, it is untried in practice and may therefore change before being published as a CAE Specification. In such a case the CAE Specification will be made as upwards-compatible as possible with the corresponding Preliminary Specification, but complete upwards-compatibility in all cases is not guaranteed.

In addition, X/Open periodically publishes:

- *Snapshots*

Snapshots are “draft” documents, which provide a mechanism for X/Open to disseminate information on its current direction and thinking to an interested audience, in advance of formal publication, with a view to soliciting feedback and comment.

A Snapshot represents the interim results of an X/Open technical activity. Although at the time of publication X/Open intends to progress the activity towards publication of an X/Open Preliminary or CAE Specification, X/Open is a consensus organisation, and makes no commitment regarding publication.

Similarly, a Snapshot does not represent any commitment by any X/Open member to make any specific products available.

X/Open Guides

X/Open Guides provide information that X/Open believes is useful in the evaluation, procurement, development or management of open systems, particularly those that are X/Open-compliant.

X/Open Guides are not normative, and should not be referenced for purposes of specifying or claiming X/Open-conformance.

This Document

Currently, word and document processing occupies two application spaces:

- personal word processing and publishing
- large-scale technical publishing.

This separation in the current marketplace leads X/Open to endorse dual systems of document interchange.

The X/Open Document Interchange Reference Model endorses several document models. X/Open-compliant document interchange assumes that the document to be interchanged conforms to one of the endorsed document models.

It is the intent of this document that an open systems user will be able to interchange documents between any two X/Open-compliant document interchange systems.

This specification relates to various international standards and recommendations which are listed in **Referenced Documents**. It should be used in conjunction with those International Standards or Recommendations.

A compliant system shall meet the definitive requirements described in this Document Interchange Format Specification.

Trademarks

UNIX[®] is a registered trademark of UNIX System Laboratories Inc. in the U.S.A. and other countries.

Palatino[®] is a registered trademark of Linotype AG and/or its subsidiaries.

X/Open[™] and the “X” device are trademarks of X Company Ltd. in the U.K. and other countries.

Referenced Documents

The following documents are referenced or are related to the content of this document:

CCITT T.4

CCITT Recommendation T.4 - Standardization of group 3 facsimile apparatus for document transmission (1988).

CCITT T.6

CCITT Recommendation T.6 - Facsimile coding schemes and coding control functions for group 4 facsimile apparatus (1988).

ISO 8613-1

ISO 8613-1 : 1989, Information processing - Text and office systems; Office Document Architecture (ODA) and interchange format - Part 1: Introduction and general principles.

ISO 8613-2

ISO 8613-2 : 1989, Information processing - Text and office systems; Office Document Architecture (ODA) and interchange format - Part 2: Document Structures.

ISO 8613-4

ISO 8613-4 : 1989, Information processing - Text and office systems; Office Document Architecture (ODA) and interchange format - Part 4: Document profile.

ISO 8613-5

ISO 8613-5 : 1989, Information processing - Text and office systems; Office Document Architecture (ODA) and interchange format - Part 5: Office document interchange format.

ISO 8613-6

ISO 8613-6 : 1989, Information processing - Text and office systems; Office Document Architecture (ODA) and interchange format - Part 6: Character content architectures.

ISO 8613-7

ISO 8613-7 : 1989, Information processing - Text and office systems; Office Document Architecture (ODA) and interchange format - Part 7: Raster graphics content architectures.

ISO 8613-8

ISO 8613-8 : 1989, Information processing - Text and office systems; Office Document Architecture (ODA) and interchange format - Part 8: Geometric graphics content architectures.

ISO 8613-1 : (to be published)

ISO 8613-1 : (to be published), Information processing - Text and office systems; Office Document Architecture (ODA) and interchange format - Part 1: DAD - A document application profile proforma and notation.

ISO/IEC 646

ISO/IEC 646 : 1991, Information technology - ISO 7-bit coded character set for information interchange.

ISO 8859-1

ISO 8859-1 : 1987, Information processing - 8-bit Single-byte coded graphic character sets - Part 1: Latin alphabet No. 1.

ISO 6937-2

ISO 6937-2 : 1983, Information processing - Coded character sets for text communication - Part 2: Latin alphabet and non-alphabetic characters.

ISO 2022

ISO 2022 : 1986, Information processing - ISO 7-bit and 8-bit coded character sets - Code extension techniques.

ISO 7350

ISO 7350 : 1984, Text communication - Registration of graphic character subrepertoires.

ISO 8824

ISO 8824 : 1987, Information processing systems - Open Systems Interconnection - Abstract Syntax Notation One (ASN.1).

ISO 8825

ISO 8825 : 1987, Information processing systems - Open Systems Interconnection - Basic encoding rules for abstract syntax notation one (ASN.1).

ISO 8632

ISO 8632 : 1987, Information processing systems - Computer graphics - Metafile for the storage and transfer of picture description information.

Part 1: Functional specification

Part 3: Binary encoding.

ISO 8879

ISO 8879 : 1986, Information processing - Text and office systems - Standard Generalized Markup Language (SGML).

ISO/IEC ISP 10610-1

ISO/IEC 10610-1, Information technology - International Standardized Profile FOD11 - Office Document Format - Simple document structure - Character content architecture only - Part 1: Document application profile.

ISO/IEC ISP 11181-1

ISO/IEC 11181-1, Information technology - International Standardized Profile FOD26 - Office Document Format - Enhanced document structure - Character, raster graphics and geometrical graphics content architectures - Part 1: Document application profile.

ISO/IEC ISP 11182-1

ISO/IEC 11182-1, Information technology - International Standardized Profile FOD36 - Office Document Format - Extended document structure - Character, raster graphics and geometric graphics content architectures - - Part 1: Document application profile.

XDIR

X/Open Document Interchange Reference Model, Snapshot, Part Number S209, 1992.

ISO 8613-7, ISO 8613-8, CCITT Recommendations T.4, T.6 and ISO 8632 are not applicable to the FOD11 document model since they concern content types not present in FOD11.

1.1 Background

Currently, word and document processing occupies two application spaces:

- **Personal word processing and publishing**

A variety of applications maintain and manipulate structured compound documents (containing text, raster graphics, and geometric graphics). The emphasis in these applications is "what-you-see- is-what-you-get" (WYSIWYG); that is, that the user's screen should as closely as possible approximate the appearance of a printed document.

- **Large-scale technical publishing**

Applications used in large-scale corporate technical publishing are markup-based, rather than WYSIWYG. The source form of a technical document is expressed in a markup language, such as the Standard Generalized Markup Language (SGML).

Procedures for manipulating SGML generally do not address the appearance of the printed document. Compositions in SGML may describe not only the structure of the information as a document, but also other abstract relationships among the information. Use of a markup language thus sometimes enables automation of additional document generation and conversion tasks.

This separation in the current marketplace leads X/Open to endorse dual systems of document interchange.

1.2 Role of Reference Model

The X/Open Document Interchange Reference Model (reference **XDIR**) endorses several document models. X/Open-compliant document interchange assumes that the document conforms to one of the endorsed document models.

This requirement is crucial. For example:

- Endorsement of ODA (Open Document Architecture¹) as a method of document interchange is insufficient, since ODA is a framework standard. It is necessary instead to endorse ODA in conjunction with one or more of the ODA-based document models that are internationally standardized (defined in the ODA International Standardized Profiles FOD11, FOD26, or FOD36 - see **Referenced Documents**) that X/Open endorses in its Document Interchange Reference Model.

1. "Open Document Architecture" was formerly called "Office Document Architecture" in the 1989 version of ISO 8613. The proposed ISO 8613:1992 will be renamed "Open Document Architecture".

- Endorsement of SGML as a method of document interchange is also insufficient. SGML is simply a language; the endorsement must include not only the syntax of SGML but also a published set of conventions for the use of SGML, in order to ensure consistent processing of the document by both participants in the interchange.

2.1 The ODA Standard

X/Open-compliant systems support the document interchange format specified in ODA (ISO 8613), for one of the FOD11, FOD26, or FOD36 document models, as one document interchange format for use among word processing and personal publishing applications.

The ODA standard describes how to address interworking of document processing applications, including character-text systems as well as systems including character text, raster graphics and geometric (vector) graphics. The standard provides a comprehensive level of features for the transfer of documents between these systems.

ODA is an International Standard in seven parts and depends on the use of other International Standards and Recommendations for the content types (see **Referenced Documents**).

ODA allows for extensions, including typographical features, colour, spreadsheets and additional types of content such as sound. It uses a mathematical formal specification language, which resolves many questions of interpretation.

ODA provides for the transfer of documents in a format based on the Abstract Syntax Notation 1 (ASN.1) and called Open Document Interchange Format (ODIF), or in a format based on the Standard Generalized Markup Language (SGML) and called Open Document Language (ODL). Both are defined in ISO 8613. Chapter 3 on page 7.

2.2 Definitions

ODA (ISO 8613) applies to the interchange of documents. ODA defines interchange as the process of providing a document to a receiving person or device, by means of data communication or by exchange of storage media. It defines documents as items such as memoranda, letters, invoices, forms and reports, which may include pictures and tabular material. The content elements used within any document may include graphic characters, geometric graphics elements and raster graphics elements.

ODA envisages that documents will be interchanged for the following purposes:

- **Presentation** - the operation of rendering the content of a document in a form perceptible to a human being. Typical presentation media are paper and video screens.
- **Processing** - includes editing and reformatting. Editing means creation and amendment of a document's structure and content. Formatting means operations that determine the layout of a document - that is, the appearance of its content on a presentation medium.

ODA allows documents to be transferred in each of the following forms:

- **Formatted form** - allowing presentation of the document (also called *final form* or *image form*)
- **Processable form** - allowing processing of the document (also called *revisable form*)
- **Formatted processable form** - allowing both presentation and processing of the document.

Each of these forms lets the originator express intentions regarding the structure and format of the document.

The interchange format of a document is convertible from a processable form to the other two forms.

The interchange format is independent of the processes used to create, edit or reproduce documents. It is also independent of the transfer medium.

ODA aims to effect document interchange so as to communicate most accurately the intentions of a document originator with respect to editing, formatting, and presentation.

2.3 Document Structure

A document may have two types of structure:

- **Logical structure** - dividing the document, based on meaning, into logical objects such as chapters, sections, figures, and paragraphs;
- **Layout structure** - dividing the document, based on layout, into layout objects such as pages and blocks.

These are alternative but complementary views of the same document. For example, a book can be regarded as consisting of chapters containing figures and paragraphs, and at the same time, as consisting of pages that contain text blocks and graphic blocks.

Typically the author specifies the logical structure, whereas typically a formatting process controls the layout structure. However, the author may specify attributes called layout directives that guide or constrain the formatting process.

2.4 Compliance

A data stream that represents a document complies with this specification if:

- It is encoded in accordance with the ASN.1 encoding rules defined in ISO 8825
- The document is constructed in accordance with the ODA abstract document architecture
- The document is encoded in accordance with the encoding and structuring rules for the document application profile corresponding to the selected open document model (that is, currently one of the International Standardized Profiles FOD11, FOD26, FOD36 - see **Referenced Documents**)
- The document is structured in accordance with the ODA characteristics and constraints (specified in clause 6 of the relevant ODA International Standardized Profile definition).

A compliant implementation is categorized as either an originator, a recipient or an originator/recipient class of implementation.

Compliant recipient and originator/recipient classes of implementation should be able to receive any data streams that comply with this specification.

Compliant originator and originator/recipient classes of implementation can be configured so as to only generate data streams that comply with this specification.

2.5 Limitations

ODA did not originally support colour. However, comprehensive support for colour has been ratified by ISO and now is being formally added to the ODA standard.

ODA currently does not support external references or *live links*, and does not support the processable characteristics of business graphics and spreadsheets. X/Open encourages enhancements to ODA in these directions.

3.1 The SGML Standard

X/Open-compliant systems support SGML document interchange utilizing any Document Type Definition (DTD) that can be specified in SGML. This provides a degree of flexibility in complex document structuring. SGML documents (marked up source instances) and declarations are all in *clear text* but SDIF (an application of ASN.1) provides a means for bundling all the parts of an SGML document (which does not have to exist as a single unit, but may exist as multiple entities/fragments that may be stored in a database) into a *unit* for interchange.

SGML/SDIF can be used to represent ODA documents where the meaning of all attributes and tags is provided by the ODA standard (with additional semantics provided by the DAPs). X/Open specifically does not prejudge the use of SGML, i.e. which DTD is to be chosen.

3.2 Definitions

- **Element** - a component of the hierarchical structure defined by a document type definition. It is identified in a document instance by descriptive markup, usually a start-tag and end-tag.
- **Document Type** - a class of documents having similar characteristics; for example, journal, article, technical manual, or memo.
- **DTD** - Document Type Definition: rules, determined by an application (of SGML), that apply SGML to the markup of documents of a particular type.

A DTD is a formal definition of the vocabulary of descriptive markup that may be used to mark up the class of documents. This includes the definition of the elements and element structures.

- **Generic Identifier (GI)** - a name that identifies the element type of an element.
- **Document Instance (Source Instance)** - instance of a document type; the portion of an SGML document that contains the elements and their content.
- **Entity (Entity management)** - a collection of characters (or data) that can be referenced as a unit. This may coincide with a storage unit.
- **SDIF-SGML Document Interchange Format** - an ASN.1 application that defines a binary wrapper that may be used when interchanging the various constituents (entities) of an SGML document.

3.3 Document Structure

SGML applications normally convey only logical structure, but SGML through its CONCUR feature can represent more than one structure (as is required for the ODL representation of an ODA formatted processable document). Most current SGML implementations deal with only the logical structure - from which formatters may derive the formatted document for display or printing.

3.4 Compliance

X/Open compliance with regard to use of SGML/SDIF is as defined in the SGML and SDIF standards. If the SGML application is that of ODL, then semantics of ODA are inherited from that standard (and its associated DAPs). The X/Open goal for SGML/SDIF document interchange is that of data consistency and understanding at the data representation level. An SGML parser is required to check DTD consistency and the marked up instance of an SGML document. Parsers, formatters and other processors of SGML documents should comply with existing X/Open guidelines to achieve application portability, but actual system reaction to data represented in an SGML document that is interchanged is not a subject of this recommendation. Such semantics are described in numerous SGML *user* standards (such as the CALS specifications) and may be endorsed by X/Open at a future date, if considered appropriate by X/Open members.

3.5 Limitations

SGML does not standardize processes which may be viewed as a limitation. However this may also suit some uses, as it allows applications to add value within the framework of SGML processing.

Effects on Implementations

4.1 Publication Requirements

It is the intent of this specification that an open systems user is able to interchange documents between any two X/Open-compliant document interchange systems.

The difference between the word processing/publishing application space and the large-scale technical publishing application space is clear. However, each X/Open-compliant implementation must publish, for the purpose of interchange format endorsement, which space it occupies - that is, it must identify whether it is a word processing/publishing application or a large-scale technical publishing application.

Alternatively it may specify that it is capable, at the user's option, of sending and receiving documents using both of the interchange formats endorsed.

4.2 Non-requirements

X/Open explicitly allows implementations to use interchange formats other than those specified in this document.

Also, X/Open allows implementations to use a format endorsed in this specification for one application space, in conjunction with applications in other application spaces.

4.3 Testing Compliance

It is straightforward to test an interchanged document (in stored form or across another transfer medium) for compliance. Compliance indicates that the form of the document is valid for interchange.

However, it is more important to test implementations for their ability to originate and receive compliant documents:

- Compliance in receiving a document means the implementation accepts as valid, and recognises, each object and element allowed in a document under the given document model. Compliance in receiving does not prove that the implementation will subsequently give the object appropriate processing, or that the information the author placed in the object has been transferred with full fidelity.
- Compliance in originating a document is a limited test that indicates that the implementation has included in the document all objects and elements that are mandatory, and did not include in the document any object or element that is prohibited, under the given document model. An implementation is not required to be able to generate all the objects or elements that a given model might allow. Compliance in sending does not prove that the implementation chose those transfer objects and elements that would best reflect the author's intentions.

Proof of compliance is not necessarily proof of a good implementation. X/Open characteristically gives implementations wide latitude to add value.

API

Application Programming Interface

attribute

A defined property of an entity or object, usually defining a characteristic feature. An attribute is named, and is often implemented as fields with values.

CALS

Computer-aided Acquisition and Logistics Support. A US Department of Defense information technology scheme to ensure open systems facility to interchange information between computer systems, for the purpose of integrating all aspects of procurement, manufacturing and maintenance information.

CCITT

Comite Consultatif Internationale de Telegraphique et Telephonique. A United Nations agency which makes recommendations regarding telecommunications. Its technical recommendations often become internationally recognised standards (V-series, X-series).

EDI

Electronic Data Interchange. The electronic exchange of structured information between computer systems.

EDIFACT

EDI for Administration, Commerce and Transport

ISO

International Organization for Standardization. Body made up of committees of representatives from the standards bodies of member countries, with particular interest in electronic interchange of information.

ISP

ISO Internationally Standardized Profile

Object-oriented

Software approach in which everything (processes, files, etc.) is represented as objects. Objects are data structures in memory that may be manipulated by the system; communication between objects is achieved by passing messages. In this regard, an object is a package of information and a description of its manipulation, and a message is a specification of one of an object's manipulations.

ODA

Office Document Architecture. More generally called Open Document Architecture, it is an architecture of documents which is defined in a set of ISO standards (8613), aimed at enabling interchange of information between computer systems.

SGML

Standard Generalized Markup Language. A family of ISO standards for labeling electronic versions of text, enabling both sender and receiver of the text to identify its structure.

