A Survey on Design Criteria for Interchange Formats

Jan Mendling
Department of Information Systems, New Media Lab
Vienna University of Economics and BA
jan.mendling@wu-wien.ac.at

Abstract: This paper describes interchange formats as an important subject of standardization. Building on reported experiences of interchange format design in specific domains, a list of design criteria is gathered addressing both the conception of the metamodel behind the interchange format and the serial representation.

1 Introduction

The standardization of XML and related specifications has triggered the definition of various XML-based interchange formats. In essence, an interchange format defines the structure of a file via a grammar or a schema that represents data relevant for a certain application domain. Independent software components can then consume data files that other applications have produced. As a consequence, a standardized interchange format provides for a simple integration of applications (see e.g. [HW04]).

There is numerous work available on the design of interchange formats, but almost every contribution focuses on the specific domain for which the interchange format has been designed. Domain-specific interchange formats have been compared for example in the area of software reengineering [WOS00] or Business Process Management [NM04]. Yet, there is a need to discuss interchange format design from a more general perspective, because design experiences in one specific domain might be valuable for designers that work on interchange format specifications in a different domain. Analyzing reports on design experiences can be helpful to establish general quality criteria. These might provide the basis to assess the suitability of an interchange format proposal for standardization no matter for which domain it has been defined.

This paper aims to contribute to a general view of interchange format design by condensing experiences made in different domains. Although the selection of papers is not exhaustive, it shows that the challenges are quite similar across different domains. The rest of the paper is structured as follows. Section 2 presents the goals and application scenarios that motivate the definition of an interchange format. Section 3 discusses quality criteria for the metamodel behind the interchange format. This aspect needs to be distinguished from quality criteria for serial representation of an interchange format which are presented in Section 4. Section 5 concludes the paper and gives an outlook on future research.

2 The Rationale of Interchange Formats

The need for an interchange format is usually identified by a group of people that work on the same or related subjects and who want to share data across incompatible software components. Basically, the resulting specification of an interchange format creates three effects: a pragmatic, an economic, and a conceptual effect.

The *pragmatic effect* of an interchange format establishes inter-operability between incompatible applications and platforms (see e.g. [BCV96], [SW99], [PLGM98], [Ko92], [Ha92], [GSM99], or [MBE95]). This leverages collaboration between different groups of people [OMGM+98], because data can be reused avoiding time-consuming re-entering [BCV96], [GSM99]. By this means data exchange via standardized interchange formats provides for a cost-efficient integration of tools [SW99], [Ko92], [GSM99] because equipping existing components with import and export capabilities requires only little programming effort. Furthermore, interchange formats can serve as intermediary formats. In this setting proprietary data of n applications is mapped to the intermediary format instead of developing $n * (n-1)$ bilateral translations (see e.g. [WHB02], [PLGM98], or [GSM99]).

Beyond inter-operability the availability of an interchange format has *economic impacts*. Its standardization grants free competition between software vendors [Cr84]. It might even create a market when it leverages the application of a previously unstandardized technology [Ko92], [MBE95]. Moreover, it encourages the development of new tools [GSM99] and enables users to choose the best tool for a particular task instead of being lock-in with a certain vendor [Cr84]. Furthermore, interchange formats have an influence on the *concepts* of the specific domain they address. They can be regarded as reference models which leverage the explication of concepts and consolidation of terminology [OMGM+98], [Eu86]. The conceptualization of the domain might employ techniques like metamodelling. Design criteria for such metamodels are discussed in the following section.

3 Metamodel behind the Interchange Format

The domain for which the interchange format is to be defined needs to be formally described. Its concepts and the relationships between them have been explicated via a metamodel or an ontology. This metamodel has to meet different design criteria including of simplicity, completeness, generality, unambiguity, and extensibility.

- *Simplicity* refers to freedom of complexity [SDSK00] in order to provide a compact metamodel that is easy to understand [MM00].
- On the other hand, *completeness* demands that a sufficient set of concepts is included [NTD98], [SDSK00] in order to provide the expressive power for representing all relevant aspects of the domain [BCV96], [Cr84], [SW99].
- Furthermore, *generality* has to be offered by the interchange format to be applicable in all scenarios that are relevant to the domain [PLGM98], [BCV96], [SDSK00],

[MM00], [Cr84], [KCE00], [Eu86]. Especially those concepts should be taken into account that are included in existing tools [SW99], [Eu86].

- Moreover, the interchange format has to offer an *unambiguous* view on the domain. Precise terms need to be chosen and related semantics have to be defined formally [BCV96]. By this means an interchange format might prove valuable for the consolidation of terminology in the respective domain [OMGM+98], [Eu86].
- *Extensibility* belongs to the most frequently mentioned criteria [BCV96], [Cr84], [SW99], [Ko92], [PLGM98], [Ha92], [Eu86], [NTD98], [SDSK00], [Eu86]. It provides for the inclusion of additional information in a predefined way. This is especially desirable, because future developments, new requirements, and changing technology might motivate unanticipated revisions of the format in a priori unknown directions. Extensibility grants a smooth integration of such new aspects.

4 Serial Representation of the Interchange Format

The domain-specific metamodel for which the above mentioned design criteria have been defined needs to be transformed to a serial representation. In order to provide for a formal specification grammars or schemas are commonly used. The design of the serial format should meet different criteria including identity to the metamodel, readability, ease of implementation, standards support, platform independence, efficiency, and free availability.

- *Identity* of metamodel and serial representation is important in order to avoid loss of information [SDSK00]. Formally, this implies that isomorphic mappings between them must be available. This criterion governs the choice for a tree- or graph-based serialization [GF92].
- *Readability* is frequently mentioned as an important feature of an interchange format [SDSK00], [KCE00], [GF92]. For the software developer the implementation and testing of a readable interchange format is much easier. Accordingly, it is related to the following criterion *ease of implementation*.
- The *ease of implementation* contributes to the success of an interchange format [Cr84], [KCE00], [GF92], [Eu86], [NTD98], [SW99]. Availability of standard parsers and related tools facilitates fast and wide-spread adoption [GSM99], [MM00]. This criterion might partially be able to explain the success of XML in contrast to the more complex and more difficult to implement SGML.
- The *support of standards* is useful on the one hand to take advantage of existing tools and solutions [NTD98]. On the other hand, building on an available specification disburdens the designer of an interchange format from technical details. Furthermore, the use of standards can leverage the combination of the interchange format with other information [NTD98], e.g. by using XML namespaces.

- Frequently, an interchange format needs to be *platform independent* [Ko92]. Accordingly, standardized character sets like UNICODE should be used for encoding.
- *Efficiency* has been an important requirement for the definition of compact graphic formats meant for multimedia presentation on the Internet [Ko92]. In this context storage efficiency [Ko92], [KCE00], [DRB+98] and access efficiency [Ko92], [DRB+98], [SDSK00] can be distinguished. The advent of XML-based graphic formats like Scalable Vector Graphics illustrate that the efficiency criterion has lost importance during the past ten years.
- *Free availability* of the interchange format specification is crucial for the adoption, otherwise people will have difficulties developing standard-compliant software [MM00].

It is interesting to notice that most of these criteria yield positive results for an XML-based serial representation: XML can be tailored to be readable and isomorphic to a giving metamodel; it is easy to implement because various parsers are available; additional XML-related tools can be utilized like XSLT processing engines; XML is platform-independent via UNICODE support; and the specification is freely available. Yet, XML is not storage efficient. Meeting most of these criteria might partially explain the success of XML as the standard for interchange format serialization.

5 Summary

In this paper we presented general design criteria for both conception of metamodels behind interchange formats and serial representation of standard interchange formats. Basically, these criteria might serve as an evaluation framework for interchange formats. Yet, the criteria for serial representation gathered from various papers require further research. They are still too general in order to compare different proposals that are based on XML. Another area of future research is the definition of an engineering process for interchange format design which might be helpful for future standardization efforts in various domains.

References

[BCV96] Bidoit, M., Choppy, C., and Voisin, F.: Interchange Format for Inter-Operability of Tools and Translation. In: *Proceedings of the 11 Workshop on Specification of Abstract Data Types*. Lecture Notes in Computer Science. pages 102–124. 1996.

[Cr84] Crawford, J. D.: An electronic design interchange format. In: *Proceedings of the 21st Conference on Design Automation*. pages 683–685. 1984.

[DRB+98] Dolin, R. H., Rishel, W., Biron, P. V., Spinosa, J., and Mattison, J. E.: SGML and XML as Interchange Formats for HL7 Messages. In: *Proceedings of AMIA Annual Symposium*. pages 720–724. 1998.

[Eu86] Eurich, J. P.: A Tutorial Introduction to the Electronic Design Interchange Format. In: *Proceedings of the 23rd Design Automation Conference*. pages 327–333. 1986.

[GF92] Genesereth, M. R. and Fikes, R. E.: Knowledge Interchange Format Version 3.0 Reference Manual. Report Logic-92-1. Stanford University. June 1992.

[GSM99] Gravitz, P. D., Sheehan, J., and McLean, T.: Common Activities in Data Interchange Format (DIF) Development. Report. McLeod Institute of Simulation Sciences. http://www.ecst.csuchico.edu/~hla/LectureNotes/99S_177SIWPaper.pdf. 1999.

[Ha92] Hamilton, E.: JPEG File Interchange Format - Version 1.02. Specification. ISO. http://www.w3.org/Graphics/JPEG/jfif3.pdf. 1992.

[HW04] Hohpe, G. and Woolf, B.: *Enterprise Integration Patterns*. Addison Wesley. 2004.

[KCE00] Kienle, H. M., Czeranski, J., and Eisenbarth, T.: Exchange Format Bibliography. In: *Proceedings of the Workshop on Standard Exchange Format (WoSEF)*. 2000.

[Ko92] Koegel, J. F.: On the Design of Multimedia Interchange Formats. In: *Proceedings of the 3rd International Workshop on Network and Operating System Support for Digital Audio and Video*. pages 262–271. 1992.

[MBE95] Meyer-Boudnik, T. and Effelsberg, W.: MHEG Explained. *IEEE MultiMedia*. 2(1):26–38. Spring 1995.

[MM00] Martin, J. and Müller, H.: Considerations on the Syntax of a Standard Exchange Format. In: *Proceedings of the Workshop on Standard Exchange Format (WoSEF)*. 2000.

[NM04] Nüttgens, M. and Mendling, J. (eds.): *Proceedings of the 1st GI Workshop XML Interchange Formats for Business Process Management*. 2004.

[NTD98] Nierstrasz, O., Tichelaar, S., and Demeyer, S.: CDIF as the Interchange Format between Reengineering Tools. In: *Proceedings of the OOPSLA Workshop on Model Engineering, Methods and Tools Integration with CDIF, Vancouver, Canada*. 1998.

[OMGM+98] Ohno-Machado, L., Gennari, J. H., Murphy, S. N., Jain, N. L., Tu, S. W., Oliver, D. E., Pattison-Gordon, E., Greenes, R. A., Shortliffe, E. H., and Barnett, G. O.: The GuideLine Iinterchange Format. *Journal of the American Informatics Association*. 5(4):357–372. July 1998.

[PLGM98] Polyak, S. T., Lee, J., Gruninger, M., and Menzel, C.: Applying the Process Interchange Format (PIF) to a Supply Chain Process Interoperability Scenario. In: *Proceedings of the ECAI workshop on Applications of Ontologies and Problem-Solving Methods, Brighton, England*. 1998.

[SDSK00] St-Denis, G., Schauer, R., and Keller, R. K.: Selecting a Model Interchange Format - The Spool Case Study. In: *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences (HICSS-33), 4-7 January, 2000, Maui, Hawaii*. 2000.

[SW99] Smith, C. U. and Williams, L. G.: A Performance Model Interchange Format. *Journal of Systems and Software*. 49(1):63–80. December 1999.

[WHB02] Wüstner, E., Hotzel, T., and Buxmann, P.: Converting Business Documents: A Classification of Problems and Solutions using XML/XSLT. In: *Proceedings of the 4th International Workshop WECWIS*. 2002.

[WOS00] *Proceedings of the Workshop on Standard Exchange Format (WoSEF)*. 2000.