



Organisation for Economic Co-operation and Development

ENV/JM/MONO(2014)9/REV1

Unclassified

English - Or. English

17 July 2019

**ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND THE WORKING PARTY
ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY**

**GLOBALLY HARMONISED SUBMISSION AND TRANSPORT STANDARD
(GHSTS) FORMAT SPECIFICATION**

**Series on Pesticides
No. 77**

JT03449828

OECD Environment, Health and Safety Publications
Series on Pesticides
No. 77

GLOBALLY HARMONISED SUBMISSION AND TRANSPORT STANDARD (GHSTS) FORMAT SPECIFICATION

An XML-based Interchange Format for Registration Applications

Version 02.00.00

IOMC

INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS

A cooperative agreement among **FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD**

Environment Directorate
ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT
Paris 2019

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 36 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in eleven different series: **Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides; Biocides; Risk Management; Harmonisation of Regulatory Oversight in Biotechnology; Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; and Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

Also published in the Series on Pesticides: [link](#)

**For this and many other Environment,
Health and Safety publications, consult the OECD's
World Wide Web site (www.oecd.org/chemicalsafety/)**

or contact:

**OECD Environment Directorate,
Environment, Health and Safety Division**

2 rue André-Pascal

75775 Paris Cedex 16

France

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

© OECD 2019

Applications for permission to reproduce or translate all or part of this material should be made to: Head of Publications Service, RIGHTS@oecd.org, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France

FOREWORD

This document has been developed by the OECD Expert Group on the Electronic Exchange of Pesticide Data.

The Expert Group investigates the possibility of harmonising information technology used in the pesticide regulatory process and, in particular, has focused on harmonising methodologies for the electronic submission of documents to regulators using a common transport mechanism based on a harmonised global XML schema. In this context, the Expert Group developed an XML-based interchange format for registration applications to regulatory authorities – the Global Harmonised Submission Transport Standard (GHSTS). The GHSTS version 1.0 was published in 2014.

GHSTS version 2.0, described in this document, includes changes to support the use of the GHSTS for electronic packages in other regulatory domains in addition to the pesticides domain and includes improvements in the lifecycle management of documents in a series of dossier electronic submissions by a registrant for a dossier regulatory action.

This GHSTS format specification document describes the Standard on both a conceptual and a technical level.

This document is published under the responsibility of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology. Its subsidiary body, the Working Group on Pesticides, agreed that the document be declassified and made available to the public.

Table of contents

Chapter 1. Preface.....	14
1.1. GHSTS Primer.....	14
1.1.1. Scope of GHSTS.....	14
1.1.2. Background and history.....	14
1.1.3. Supporting stakeholders.....	15
1.1.4. The role of the OECD.....	15
1.2. Scope and audience of this document.....	15
1.3. Organisation of the document.....	16
1.4. Filename and version.....	16
1.5. Contacts.....	17
Chapter 2. Context of GHSTS.....	18
2.1. Usage of GHSTS in the submission process.....	18
2.2. GHSTS and different submission Types.....	20
2.2.1. Joint review in the pesticides domain.....	20
2.3. Related standards in the pesticides domain.....	21
2.3.1. Submission standards on national / regional level.....	21
2.3.2. IUCLID.....	22
2.3.3. OECD Harmonised Templates for Reporting Chemical Test Summaries (OECD Harmonised templates - OHT).....	23
2.4. GHSTS components.....	23
2.4.1. GHSTS core components.....	24
2.4.2. GHSTS additional components.....	28
2.4.3. Lifecycle and dependencies of GHSTS core components.....	29
Chapter 3. Introduction to GHSTS.....	30
3.1. GHSTS Design Principles.....	30
3.1.1. Focus on administrative metadata.....	30
3.1.2. Support for different regulated domains.....	30
3.1.3. Variation with picklists.....	30
3.1.4. Parallel viewpoints.....	31
3.1.5. Parallel navigation scenarios.....	32
3.1.6. Document-based organisation.....	32
3.1.7. Self-contained submission package.....	32
3.1.8. Integration of XML-based content.....	33
3.1.9. Table of contents as main structuring element for documents.....	33
3.1.10. Adaptiveness with respect to different standard ToC templates.....	33
3.1.11. Lifecycle management.....	33
3.1.12. Persistent identifiers of information concepts.....	34
3.1.13. Information specific for regulatory authorities.....	34
3.2. GHSTS logical concepts.....	34

3.2.1. Logical concepts and XML syntax.....	35
3.2.2. Overview	35
3.2.3. Metadata	37
3.2.4. Dossier regulatory action	40
3.2.5. Submission	40
3.2.6. Dossier.....	40
3.2.7. Table of Contents	41
3.2.8. Legal entity.....	42
3.2.9. Product	42
3.2.10. Component	43
3.2.11. File.....	44
3.2.12. File type.....	44
3.2.13. Document	45
3.2.14. Document Family	48
3.2.15. Document type	49
3.2.16. File formats	49
3.2.17. Type of content changes on file level.....	50
3.2.18. Content status changes of documents.....	51
3.2.19. Dependencies between content status changes and metadata changes for DOCUMENT .	53
3.2.20. Node assignment status	53
3.2.21. Differences between content status change and node assignment status	54
3.2.22. Full and incremental submissions	54
3.2.23. Impact of document content status on evaluation price	55
3.2.24. GHSTS Picklists.....	55
3.2.25. Regulated Domain.....	58
3.2.26. Persistent identifiers	58
3.2.27. Further identifiers on submitter and regulatory authority level.....	60
3.2.28. Information for and targeting at individual receivers.....	61
3.2.29. Internal and external references	62
Chapter 4. GHSTS schema definition	66
4.1. Overview.....	66
4.2. Filename and version	66
4.3. Scope.....	66
4.4. Namespace	67
4.5. Unique keys	67
4.6. Combination of value and value type	67
4.7. Notation conventions	68
4.7.1. Elements	68
4.7.2. Attributes.....	68
4.7.3. Occurrence indicators.....	69
4.8. Boilerplate for element description.....	69
4.8.1. Child elements.....	69
4.8.2. Attributes.....	70
4.8.3. Constraints.....	70
4.8.4. Example.....	70
4.9. Elements.....	71
4.9.1. Overview	71
4.9.2. GHSTS	71
4.9.3. REGULATED_DOMAIN.....	72

4.9.4. SENDER_ENVIRONMENT	72
4.9.5. BUILDER.....	73
4.9.6. VALIDATOR.....	73
4.9.7. RECEIVERS	74
4.9.8. RECEIVER	74
4.9.9. SENDER	75
4.9.10. PRODUCT	76
4.9.11. PRODUCT_RA.....	77
4.9.12. COMPOSITION.....	78
4.9.13. COMPONENT	78
4.9.14. DOSSIER	79
4.9.15. REFERENCED_DOSSIER	81
4.9.16. DOSSIER_RA.....	81
4.9.17. SUBMISSION.....	82
4.9.18. PRODUCT_USE.....	83
4.9.19. PRODUCT_FUNCTION	84
4.9.20. COMPONENTS.....	84
4.9.21. COMPONENT	84
4.9.22. DOCUMENTS	86
4.9.23. DOCUMENT	86
4.9.24. DOCUMENT_RA.....	87
4.9.25. OTHER_NATIONAL_GUIDELINE	89
4.9.26. RA_DOCUMENT_NUMBER.....	90
4.9.27. DOSSIER_CONTEXT.....	90
4.9.28. DOCUMENT_GENERIC.....	91
4.9.29. DOCUMENT_CONTENT_STATUS_HISTORY	96
4.9.30. REFERENCED_DOCUMENT.....	97
4.9.31. RELATED_TO_COMPONENT.....	98
4.9.32. REFERENCED_TO_FILE	99
4.9.33. FILES	100
4.9.34. FILE	100
4.9.35. FILE_RA.....	101
4.9.36. FILE_GENERIC	102
4.9.37. TOC.....	103
4.9.38. STANDARD_TOC_REFERENCE	105
4.9.39. STRUCTURE.....	105
4.9.40. LEGAL_ENTITIES	106
4.9.41. LEGAL_ENTITY	106
4.9.42. CONTACT_ADDRESS.....	107
4.9.43. CONTACT_PERSON.....	109
4.10. Basic XML types	110
4.10.1. TOC node type	110
4.10.2. Submission number type	112
4.11. Usage of external type definitions	112
Chapter 5. Referenced schema definitions.....	113
5.1. Standard ToC schema definition.....	113
5.1.1. Overview	113
5.1.2. Relation to the submission schema definition.....	113
5.1.3. Filename and version	113

5.1.4. Namespace	114
5.1.5. Unique keys.....	114
5.1.6. Notation conventions.....	114
5.1.7. Elements	114
5.1.8. Basic types.....	116
5.1.9. Usage of external type definitions.....	117
5.2. CommonComponents XSD	117
5.2.1. Overview	117
5.2.2. Usage in main XSD.....	117
5.2.3. Filename and version	118
5.2.4. Namespace	118
5.2.5. Picklist Item type.....	118
5.2.6. String definition.....	120
5.2.7. Relative file path type	120
5.2.8. Version number type	121
5.2.9. PID type	121
5.2.10. MD5 type.....	123
Chapter 6. GHSTS Package Definition	124
6.1. Overview of Folder structure	124
6.1.1. Usage of Content folder	125
6.2. Storage of multiple submissions of the same dossier	125
6.3. Folder naming conventions and constraints.....	126
6.4. File naming conventions	126
6.5. Platform and media support.....	127
Chapter 7. Using GHSTS.....	128
7.1. OECD GHSTS website.....	128
7.2. Creation and publication of a GHSTS submission	128
7.3. Technical Validation.....	129
7.4. Transport of a GHSTS submission	129
7.5. Consumption of a GHSTS submission	129
Chapter 8. Appendix	130
8.1. Glossary of Terms.....	130
8.2. Standards References.....	137
8.3. Copyright and Trademark Information	137
8.4. Version history.....	138

Tables

Table 1 – Contacts	17
Table 2 – GHSTS Release Package Matrix example.....	29
Table 3 – Values for metadata changes	38
Table 4 – Concepts that track metadata changes	40
Table 5 – Overview of types of FILE	45
Table 6 – GHSTS PDF profile.....	50
Table 7 – document content status indicator	52
Table 8 – Node assignment status values.....	54

Table 9 – full and incremental submission	55
Table 10 – Overview of GHSTS elements controlled by core picklists	57
Table 11 – Overview of elements controlled by non-core picklists	58
Table 12 – Overview of concepts identified by PID generated by the registrant	59
Table 13 – Overview of concepts identified centrally managed PID	60
Table 14 – Legal entity specific identifier	60
Table 15 – Overview of RA-specific information	61
Table 16 – Overview of references	64
Table 17 – internal references with unique keys	67
Table 18 – GHSTS elements with “value type / value” pattern	68
Table 19 – Notation and graphical representation of occurrence indicators	69
Table 20 – boilerplate sample: child elements definition	69
Table 21 - boilerplate sample: attributes definition	70
Table 22 - boilerplate sample: constraints definition	70
Table 23 –child elements definition for element GHSTS	72
Table 24 - attributes definition for element GHSTS	72
Table 25 –child elements definition for element SENDER_ENVIRONMENT	73
Table 26 - attributes definition for element BUILDER	73
Table 27 - attributes definition for element BUILDER	73
Table 28 –child elements definition for element RECEIVERS	74
Table 29 –child elements definition for element REICEIVER	75
Table 30 - attributes definition for element RECEIVER	75
Table 31 –child elements definition for element SENDER	75
Table 32 - attributes definition for element SENDER	76
Table 33 –child elements definition for element PRODUCT	77
Table 34 –child elements definition for element PRODUCT_RA	78
Table 35 - attributes definition for element PRODUCT_RA	78
Table 36 - constraints definition for element PRODUCT_RA	78
Table 37 –child elements definition for element COMPOSITION	78
Table 38 –child elements definition for element COMPOSITION	79
Table 39 - attributes definition for element COMPOSITION	79
Table 40 - constraints definition for element COMPONENT below PRODUCT	79
Table 41 –child elements definition for element DOSSIER	81
Table 42 –child elements definition for element REFERENCED_DOSSIER	81
Table 43 –child elements definition for element DOSSIER_RA	82
Table 44 - attributes definition for element DOSSIER_RA	82
Table 45 - constraints definition for element DOSSIER_RA	82
Table 46 –child elements definition for element SUBMISSION	83
Table 47 - constraints definition for element SUBMISSION	83
Table 48 –child elements definition for element COMPONENTS	84
Table 49 –child elements definition for element COMPONENT	85
Table 50 - attributes definition for element COMPONENT	86
Table 51 - constraints definition for element COMPONENT	86
Table 52 –child elements definition for element DOCUMENTS	86
Table 53 –child elements definition for element DOCUMENT	87
Table 54 - attributes definition for element DOCUMENT	87
Table 55 - constraints definition for element DOCUMENT	87
Table 56 –child elements definition for element DOCUMENT_RA	88
Table 57 - attributes definition for element DOCUMENT_RA	89
Table 58 - constraints definition for element DOCUMENT_RA	89

Table 59 –child elements definition for element OTHER_NATIONAL_GUIDELINE	89
Table 60 –child elements definition for element RA_DOCUMENT_NUMBER	90
Table 61 –child elements definition for element DOSSIER_CONTEXT	91
Table 62 –child elements definition for element DOCUMENT_GENERIC	94
Table 63 - constraints definition for element DOCUMENT_GENERIC	96
Table 64 –child elements definition for element DOCUMENT_CONTENT_STATUS_HISTORY	96
Table 65 - attributes definition for element DOCUMENT_CONTENT_STATUS_HISTORY ..	97
Table 66 - constraints definition for element DOCUMENT_CONTENT_STATUS_HISTORY	97
Table 67 –child elements definition for element REFERENCED_DOCUMENT	98
Table 68 –child elements definition for element RELATED_TO_COMPONENT	99
Table 69 - attributes definition for element RELATED_TO_COMPONENT	99
Table 70 –child elements definition for element REFERENCED_TO_FILE	99
Table 71 - attributes definition for element REFERENCED_TO_FILE	100
Table 72 –child elements definition for element FILES	100
Table 73 - constraints definition for element FILES	100
Table 74 –child elements definition for element FILE	101
Table 75 - attributes definition for element FILE	101
Table 76 –child elements definition for element FILE_RA	101
Table 77 - attributes definition for element FILE_RA	102
Table 78 - constraints definition for element FILE_RA	102
Table 79 –child elements definition for element FILE_GENERIC	103
Table 80 - constraints definition for element FILE_GENERIC	103
Table 81 –child elements definition for element TOC	104
Table 82 –child elements definition for element STANDARD_TOC_REFERENCE	105
Table 83 - constraints definition for element STANDARD_TOC_REFERENCE	105
Table 84 –child elements definition for element STRUCTURE	105
Table 85 –child elements definition for element LEGAL_ENTITIES	106
Table 86 –child elements definition for element LEGAL_ENTITY	107
Table 87 - attributes definition for element LEGAL_ENTITY	107
Table 88 –child elements definition for element CONTACT_ADDRESS	109
Table 89 - constraints definition for element CONTACT_ADDRESS	109
Table 90 –child elements definition for element CONTACT_PERSON	110
Table 91 - constraints definition for element CONTACT_PERSON	110
Table 92 –child elements definition for element TOC_NODE	111
Table 93 –child elements definition for element TOC2DOC	112
Table 94 - attributes definition for element TOC2DOC	112
Table 95 –child elements definition for element TOC	115
Table 96 –child elements definition for element STRUCTURE	116
Table 97 –child elements definition for element TOC_NODE	117
Table 98 - constraints definition for element TOC_NODE	117
Table 99 –child elements definition for type PicklistItemType	119
Table 100 –child elements definition for element DECODE	119
Table 101 - attributes definition for element DECODE	119
Table 102 - constraints definition for element PID	122
Table 103 – overview of folder structure	125
Table 104 – folder naming conventions	126
Table 105 – file naming conventions	127
Table 106 – standards references	137
Table 107 – version history	138

Figures

Figure 1 - Dossier Regulatory Action	18
Figure 2 – Submission process	19
Figure 3 –Submission package overview	20
Figure 4 – Overview of GHSTS components	24
Figure 5 – Overview of ToC references	26
Figure 6 – Support of parallel viewpoints	32
Figure 7 – Example of the metada in XML linked to a document (excerpt only)	35
Figure 8 –Main GHSTS concepts and their relations	36
Figure 9 – logical concept of FILE	44
Figure 10 – logical concept of DOCUMENT	46
Figure 11 – Document Family, Document and File	48
Figure 12 – PicklistItem Type usage	56
Figure 13 – overview of receiver-specific references	62
Figure 14 – overview of external references	65
Figure 15 – value and value type	68

Chapter 1. Preface

1.1. GHSTS Primer

1.1.1. Scope of GHSTS

The Global Harmonised Submission Transport Standard (GHSTS) is a standardized set of technical specifications used to assemble electronic files for the electronic submission of pesticide registration applications to regulatory authorities in a predefined manner. The GHSTS can also be used for other regulated products. The files contained can take any form appropriate for the business needs. Standard text processing and spreadsheet formats, PDF and XML data files are just some of the file formats which can be transferred using the GHSTS. Once the files are assembled according to the specifications, they can be transferred from one business entity to another with the receiving entity able to extract the files for use in the regulatory process. At present, the majority of the pesticide regulatory processes in most nations is centred on documents, for example the “Dossier”, “Monograph”, “study”, “DER”. The required exchange of these documents is central to the existing processes and policies.

Limited metadata are included in the GHSTS. Only enough information is included to identify who the submitting entity is, the purpose of the files contained within the GHSTS, and how to handle the files once received. The GHSTS itself is not intended to use the content of the files. It is not a standard for the information itself.

The GHSTS is not a software application. Other Information Technology (IT) systems can be constructed to utilize the standard and, by doing so, could easily transfer data to each other.

1.1.2. Background and history

With a number of disparate systems and methodologies in place around the world, multinational submissions can be both resource intensive for the industry to assemble and awkward for the authorities to review.

Currently industry has to comply with many different electronic standards for the submission of data to regulatory authorities in the pesticide registration area. Examples of such systems are the Australian Pesticides and Veterinary Medicines Authority data list, CADDY (Computer Aided Dossier and Data Supply)/CADDY-xml in the European Union, e-Index (electronic index) / e-PRS (Electronic Pesticide Regulatory System) in Canada, PRISM (Pesticide Registration Information System) in the United States and IUCLID (International Uniform Chemical Information Database) / REACH IT used in Europe for classification and labelling purposes.

Other regulatory authorities are also developing their own e-submission standards resulting in additional hurdles for companies engaged in submissions globally, particularly in Joint Reviews of an agricultural pesticide submission by two or more regulatory authorities. Often this results in incompatible deliverables for the applicant and in a communication dilemma between regulatory authorities and with the applicant. In addition, most of the information submitted for Joint Reviews has been unstructured, making efficient and effective information management very difficult.

The existence of multiple standards means that applicants must duplicate their submission preparation effort as they need to provide a list of data elements with matching dossier and document metadata such as title, author, guideline number and report number in a proprietary format for every electronic submission to a regulatory authority in support of an application for registration. In some cases regulators will provide a software tool to prepare electronic submissions, but data entry of metadata is usually manual work for the applicant. Manual compilation of large dossiers is an inefficient and error prone process.

The OECD engaged with efforts to harmonise the pesticide review process world wide. To further the efforts of global harmonisation, the OECD Expert Group on the Electronic Exchange of Pesticide Data was tasked with investigating the possibility of harmonisation in the information technology used in the pesticide regulatory process, where that harmonisation might occur, and what would be required for the harmonisation to take place. As part of its investigation, the Expert Group created the Transport Mechanism Subgroup which looked specifically at the ability to harmonise the various methods used to submit information to the regulatory authorities.

These two groups focused their attention on the development of a common method of electronic submission – the Global Harmonised Submission Transport Standard (GHSTS).

GHSTS is a generalised standard which focuses primarily on administrative data and document life cycle management. The standard can be applied to a broad range of regulated areas which require a scientific/risk assessment. However specific terminologies and requirements are defined for different regulatory domains. For this reason the concept of “regulated domain” is defined (e.g. pesticides, genetically modified organisms, feed additives). The “regulated domain” defines a subset of the terms in the GHSTS picklists which are applicable for electronic submissions transmitted for evaluation according to the defined “regulated domain”. For further details about the regulated domain concept, please refer to chapter 3.1.2.

1.1.3. Supporting stakeholders

Participants in the OECD Expert Group on the Electronic Exchange of Pesticide Data and the Transport Mechanism Subgroup include members from OECD countries and the European Commission represented by the European Chemicals Agency, the European Food Safety Authority and experts from the pesticide industry.

1.1.4. The role of the OECD

Ownership of the GHSTS lies with OECD. The responsibility for monitoring updates of the GHSTS lie with the OECD Expert Group on the Electronic Exchange of Pesticide Data, under the supervision of the OECD Working Group on Pesticides.

All core components of the GHSTS, as well as some information about additional GHSTS components, are made available on the OECD public website (see chapter 7.1).

1.2. Scope and audience of this document

The intended audiences of this document are as follows:

- Responsible stakeholders, at a business project level in organisations of registrants and authorities, who are considering adoption of the GHSTS and need to know details about GHSTS, its usage and main characteristics.

- IT project leads from IT departments responsible for guiding development projects supporting GHSTS, such as publication components or ingestion components, who need to evaluate the impact of adopting GHSTS for their in-house dossier management systems.
- Business domain experts who are in charge of analysing generation, validation or import errors.
- IT consultants in charge of GHSTS-related development who require a detailed specification of GHSTS. Expertise in XML and XSD technology is assumed, general knowledge of document management is helpful.

To support the needs of these different audiences, the document is divided in different sections that describe GHSTS on a conceptual level and then provide technical in-depth details.

1.3. Organisation of the document

The Specification is organized in the following chapters:

- Chapter Chapter 2. “Context of GHSTS” sets GHSTS in the context of the submission process and other related standards. An overview of GHSTS components is presented. Scenarios for the transition to GHSTS are outlined.
- Chapter Chapter 3. “Introduction to GHSTS” describes the design principles behind GHSTS and explains the basic concepts in GHSTS on a logical level, not obscured by technical details.
- Chapter Chapter 4. “GHSTS schema definition” is the technical reference for the structure of the GHSTS XML backbone file that is the core of a submission using GHSTS.
- Chapter Chapter 5. “Referenced schema definitions” describes the technical references from the GHSTS XML backbone file to two auxiliary XSDs (GHSTS Table of Contents [ToC], CommonComponents XSD).
- Chapter Chapter 6. “GHSTS Package Definition” specifies the structure and characteristics of the GHSTS package in which all content and additional files are packed for transport.
- Chapter Chapter 7. “Using GHSTS” gives some guidance on how GHSTS ought to be used, from creation via testing to viewing.
- Chapter Chapter 8. “Appendix” contains a number of appendices, such as a glossary of terms. It also contains the version history where the changes between the versions of this specification are detailed.

This document can be read in its entirety or in parts as a reference manual using the cross-references to navigate through the document as needed.

1.4. Filename and version

The GHSTS format specification (this document) is versioned using a version numbering scheme which is composed of a two digit major, a two digit minor and a two digit patch level number separated by single dots (XX.XX.XX). Major versions are used for global functional changes, minor versions for local changes and patch level versions for error corrections. Patch level versions do not require new versions of other GHSTS components (see chapter 2.4).

The filename of the GHSTS format specification uses the same versioning scheme with the hyphen instead of the dot: The file naming convention of the GHSTS format specification is:

ghsts_XX-XX-XX.<format suffix>

for example

ghsts_02-00-00.pdf

1.5. Contacts

Any entity needing assistance regarding the GHSTS should contact the OECD at <mailto:ehscont@oecd.org>. The OECD Secretariat will then direct questions to the OECD body responsible for the maintenance of the GHSTS.

Role	Name	Email
Organisation responsible for the GHSTS	OECD Secretariat	ehscont@oecd.org

Table 1 – Contacts

Chapter 2. Context of GHSTS

2.1. Usage of GHSTS in the submission process

The GHSTS standard supports applicants and regulatory authorities in the course of a dossier regulatory action for a product registration. This chapter explains the context in which GHSTS is applied and gives an overview of the process. Technical details are explained in subsequent chapters.

A dossier regulatory action is a series of events accomplished by a regulatory authority beginning with the submission of an application with data by a registrant and ending with a unique regulatory decision. The following figure shows a sample for one dossier regulatory action, for a given active ingredient or product.

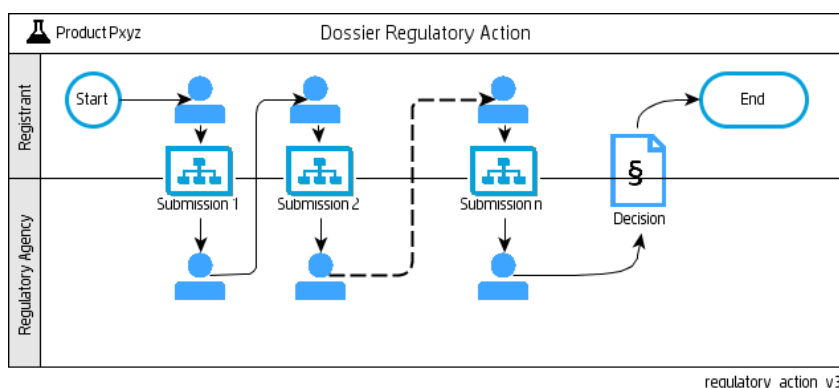


Figure 1 - Dossier Regulatory Action

As illustrated, the registrant initiates the dossier regulatory action with the initial submission. The submission is verified by the regulatory authorities, eventually leading to new submissions to fulfil the regulatory requirements. Finally, the regulatory authority submits the decision for the application to the registrant. This ends this specific regulatory action.

The dossier is the set of documents prepared and continuously maintained by the applicant throughout the lifecycle of a regulatory action. A dossier has at least one submission. A submission is the compilation of documents in a structured form according to the given regulatory requirements. Multiple submissions can be submitted for each dossier, they are differentiated with version numbers.

GHSTS supports the dossier regulatory action by defining a standard for the representation of the individual submissions in a regulatory action. The representation is called a GHSTS package and contains all required information for a submission as well as lifecycle information to set the submission in the correct context of the dossier regulatory action.

GHSTS also supports the joint submission and the joint review scenario with multiple senders (registrants) and multiple receivers (regulatory authorities) for the same regulatory action.

The following figure provides details about one single submission step. For information on the availability of the components depicted below, see chapter 2.4. The domain of the registrant is on the left

side of the figure, the transport domain in the middle and the regulatory authority domain on the right side.

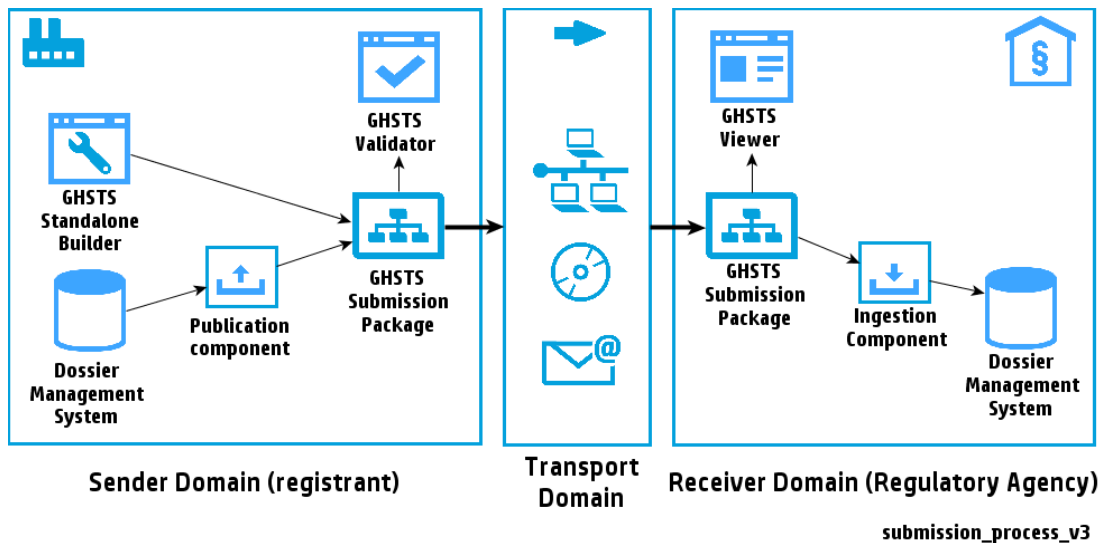


Figure 2 – Submission process

A GHSTS submission package can be created in one of the following ways:

- The required information and content for the submission is kept in a dossier management system and the GHSTS submission package is created by a publication component that generates the GHSTS submission package format.
- A GHSTS standalone builder is used that is capable of creating GHSTS submission packages based on information directly supplied to the standalone builder.

After the creation of the GHSTS submission package it must be validated using the technical validation ruleset (see chapter 2.4.1.10). As well, additional sender and / or recipient-specific validation rules may be applied.

The validated package can then be submitted to the regulatory authority as receiver. GHSTS does not restrict the parties on how a package should be transported, for example network protocols, offline media like CD/DVD, USB and hard-drives, or email may be used.

A submission package consists of two main parts:

- The GHSTS XML backbone contains the metadata of the submission as well as references to the submission documents. The GHSTS XML main backbone can also contain external references (see chapter 3.2.29).
- A standalone browser-based GHSTS HTML representation of the submission in simple HTML format is included in the submission package to allow the static human-readable display of the most relevant information via a web browser.

The receiver can work with the GHSTS package in at least the following ways:

- A GHSTS viewer can be used to open and view the submission package. A GHSTS viewer is not part of the submission package. It must be downloaded and installed once and then can be used for all GHSTS dossiers. A GHSTS viewer will use the information in the GHSTS XML backbone.

- The HTML representation of the information can be used for viewing the package in a simple way when a GHSTS viewer with more advanced functionality is not available. This HTML file also serves the purpose of long-term human-readable visualization.
- A GHSTS ingestion component can parse the package and ingest the data into an electronic dossier management system for review and approval. After ingestion, the package can be kept for auditing reasons. Subsequent activities will be performed with processes and tools outside of GHSTS.

If the submission is not the initial submission of the regulatory action then it will likely contain external references to predecessor packages.

- During ingestion into a document management system these external references need to be resolved.
- A GHSTS viewer and the HTML representation can resolve references to predecessor submissions if all packages of the same regulatory action are placed in one common file directory.

The figure below shows a high-level overview of the internals of submission packages.

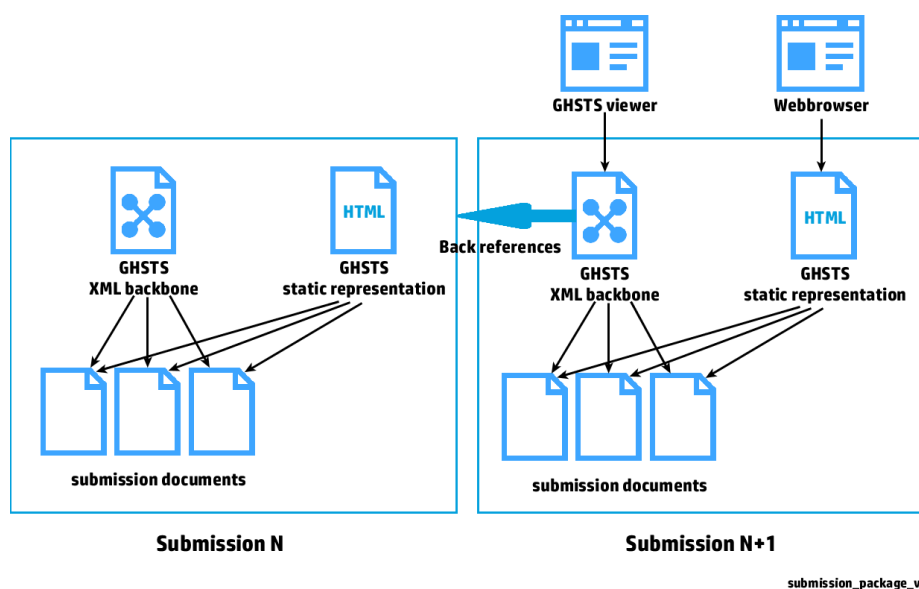


Figure 3 –Submission package overview

2.2. GHSTS and different submission Types

2.2.1. Joint review in the pesticides domain

The regulatory pesticide authorities of two or more countries may share the evaluation work in a “Joint Review” of an agricultural pesticide submission (dossier) for a new component (e.g. in pesticides an active ingredient and/or associated end-use product(s) or an additional use for an existing active ingredient or end-use product.) The participating regulatory authorities review the work of the primary reviewers for each particular science discipline, and the end product (ideally a complete monograph or key components of the monograph) is used by all participating countries (and others) as the basis for regulatory decisions.

A Joint Review requires that the pesticide dossier be submitted to all participating regulatory authorities simultaneously. This is a formal process in which timelines and work allocation are negotiated in advance. Data reviews are exchanged and peer reviewed, and there is agreement on both the documentation to be produced and the decision-making target date, i.e. the date on which the decision (which may be a proposed decision) is communicated to the applicant.

Joint review projects are not managed by the OECD Secretariat or any official OECD body. Rather, they are coordinated by various groups of countries (which may or may not all be OECD members) and pesticide companies. The OECD encourages regulatory authorities to adopt the use of the GHSTS and the OECD numbering system for pesticide dossiers in order to allow for easier communication among regulatory authorities during Joint Reviews and to reduce the regulatory burden on applicants.

2.3. Related standards in the pesticides domain

2.3.1. Submission standards on national / regional level

As a first step in the development of the GHSTS, the OECD Expert Group on the Electronic Exchange of Pesticide Data focused on the metadata used in each of the electronic standards already in place to submit information to regulatory authorities and developed a single set of metadata that could be used by all authorities. This set of metadata includes elements common to all the existing systems along with additional elements required for specific systems. Below are the specific electronic submission standards on national and regional levels whose metadata were reviewed.

2.3.1.1. CADDY

CADDY is the first electronic submission format for pesticides. It has been developed jointly by regulatory authorities in the European Union and industry since 1995 and is currently the e-submission standard used in the European Union for pesticides. The strategic goal behind CADDY was to facilitate the provision of dossiers for pesticides to regulatory authorities, the long-term archiving of such dossiers, the accessibility of information contained in such dossiers and the examination and assessment of dossiers by regulatory authorities in a cost-effective manner using electronic media.

The initial version of CADDY focused on replacing paper submissions and used an image-based page standard (TIFF) and associated metadata to allow easy searching, retrieving viewing and printing of documents within a submission. Later the development focus shifted towards facilitating the review of submissions by regulators and additional functionality to this effect was added. In 2005 the CADDY specification was migrated to an xml and PDF/A platform, while maintaining backward compatibility of the dossier metadata and providing a migration path to easily convert existing CADDY dossiers to the current CADDY-xml standard 03.07.00.

2.3.1.2. e-PRS/e-Index Builder

Electronic Pesticide Regulatory System (e-PRS) of the Health Canada Pest Management Regulatory Agency (PMRA), implemented in 1998, is an integrated electronic database designed to enable the Agency to manage regulatory submissions and associated information related to pest control products and active ingredients. The e-Index Builder is a standalone software application which facilitates the creation by an applicant of an electronic index (in XML format) that fully describes each document to be submitted in support of an application to register or amend a pest control product registration. The electronic index created by the applicant with the tool includes information on the documents' metadata and the context in which it is being submitted. The e-Index builder supports data coding systems of the OECD, PMRA and United States Environmental Protection Agency (US EPA).

Once indexed, applicants may either attach an electronic document, mail in the paper version of the document, or “cross-reference” the document to one that is already in the PMRA’s ePRS database. When the index is completed, the applicant submits it to the Agency as a “PMRA Regulatory ZIP” (PRZ) file. This file includes the e-Index as well as any attached electronic documents.

In summary, the e-Index Builder is an indexing and compiling tool for use by applicants in the preparation of their regulatory submissions, whereas the e-PRS is a database which enables the PMRA to manage and view the many documents associated with the products and active ingredients it regulates.

2.3.1.3.e-PRISM

In May 2003, the United States Environmental Protection Agency (US EPA) introduced an electronic submission program where registrants would submit electronic copies of studies after paper copies were submitted and processed for formatting requirements (PR notice 86-5). The intent of the electronic copies was to assist in the review process allowing the scientist to search for content within the study and extract that content for use in the Agency’s evaluation, even though the paper copies were considered the official copy of the study.

In July 2008, the United States introduced a revised method for electronic submission that is more advanced and consistent with current technology standards. The electronic submission process, e-PRISM (now e-Dossier), is based on the Canadian e-Index and uses much of the same technical specifications. Even though the U.S. e-PRISM methodology is based on the Canadian approach, some differences still exist.

2.3.2. IUCLID

IUCLID (International Uniform Chemical Information Database) is a software tool used to capture, store, maintain, and exchange data on intrinsic and hazard properties of chemical substances according to the format of the OECD Harmonised Templates for Reporting Chemical Test Summaries (OHT). Distributed free of charge, the software is maintained by the European Chemicals Agency (ECHA).

The software is especially useful to chemical industry companies and to government authorities. IUCLID was built to comply with the European Regulation on Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), a European Union law covering the production and use of chemical substances and was recommended as a tool for elaborating dossiers in the context of the OECD Cooperative Chemicals Assessment Programme. It is now also used in support of the submission requirements of the European Regulation on the classification, labelling and packaging of substances and mixtures (CLP Regulation) and the European Union Biocidal Product Regulation. IUCLID 6, released in 2016, offers customisation possibilities for use in different jurisdictions and is being used to support European Poison Centres notifications and Australian Industrial Chemicals assessment and is promoted for developments in international harmonisation.

Data that can be stored and maintained with IUCLID include information about the chemical substance and its composition, reference information like substance identifiers, classification and labelling, use and exposure, physical and chemical properties, toxicological properties, eco-toxicological properties and the modification history. The IUCLID data model also features Biocides/Pesticides elements.

While the majority of pesticide authorities’ processes are built around unstructured data and documents, the XML of the OECD Harmonised Templates (OHT) for Reporting Chemical Test Summaries contained in IUCLID are used to present structured data. Study summaries in the OECD XML Harmonised Template format may be exported from, exchanged and imported to IUCLID in individual XML files.

2.3.3. OECD Harmonised Templates for Reporting Chemical Test Summaries (OECD Harmonised templates - OHT)

Beginning in 2002 the OECD undertook an effort to harmonise, where possible by the use of templates and XML tags, tools for submission, evaluation and exchange of chemical data for the regulation of new and existing industrial chemicals, agricultural pesticides and biocides. This led to the establishment of the OECD Expert Group on Harmonising Templates and eventually, the development of standard data formats consisting of end point templates and XML schema for reporting studies done on chemicals to determine their properties or effects on human health and the environment (e.g., hydrolysis, skin irritation, repeat dose toxicity, etc.). These templates can be used for reporting summary results for testing on any type of a chemical (e.g., industrial chemicals, pesticides, biocides).

The Templates are aimed at developers of database systems, as they prescribe the formats by which information can be entered into and maintained in database. The XML schema are technical specifications that define the data structure and relation of that data.

The OECD XML study templates are presently used mainly by the IUCLID software, and the OECD QSAR Toolbox for filling gaps in (eco) toxicity data needed for assessing the hazards of chemicals, and subsets of template field formats are implemented in the eChemPortal - The Global Portal to Information on Chemical Substances of the OECD, and the Metabolism Pathways database (tool developed by the United States).

As the GHSTS standard is content-agnostic, content from IUCLID can be added as files in a GHSTS submission package, for example, the OHT in their XML format as defined by the XSD published by the OECD.

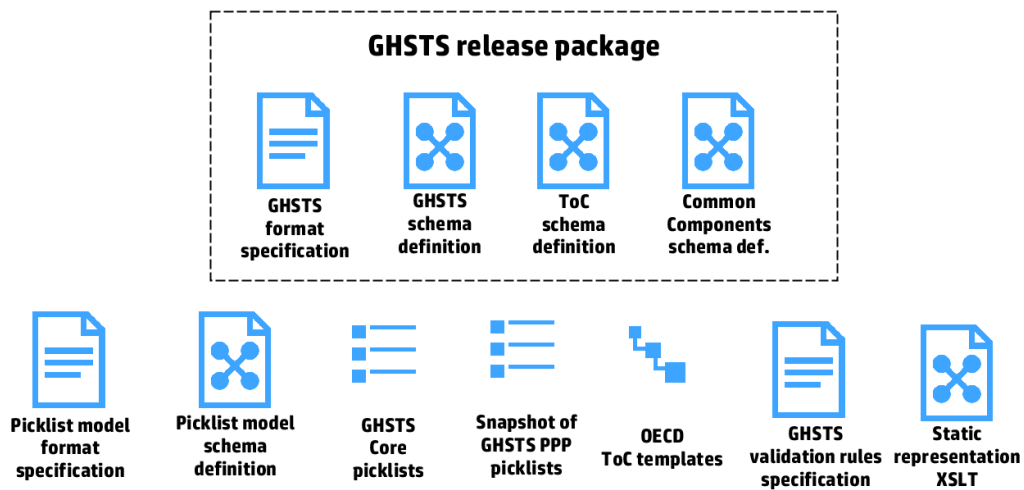
2.4. GHSTS components

This chapter provides an overview of GHSTS existing components that make up the Standard as well as potential GHSTS components that would ideally support the use of GHSTS. The components can be divided into two classes:

- **GHSTS Core components:** These are components, that make up the Standard and other supporting elements, controlled and managed by OECD member countries in the context of the OECD in order to achieve international harmonisation. Part of the core components form the “GHSTS release package”, which are the central GHSTS core components that are always published together (see chapter 2.4.3).
- **Additional components:** Components that support the use of GHSTS but are not managed by the OECD.

The following figure shows the core components in the upper rows and additional components in the bottom row.

OECD core components supported by OECD member countries



Additional components that support the use of GHSTS but are not supported by the OECD



overview_elements_v7

Figure 4 – Overview of GHSTS components

This chapter describes the core components and the additional components. For current information about the status of GHSTS components please refer to the OECD GHSTS website (see chapter 7.1).

2.4.1. GHSTS core components

2.4.1.1. GHSTS format specification

The GHSTS format specification (or short “format specification”) is the current document. The format specification focuses specifically on the GHSTS Schema Definition, the GHSTS Common Components Schema Definition and the GHSTS Table of Contents Schema Definition.

2.4.1.2. GHSTS Schema Definition

The GHSTS Schema Definition (GHSTS XSD) contains the technical definition of the GHSTS XML backbone file. It is generated from information contained in the GHSTS format specification and supplied in the XML Schema Definition format (XSD).

2.4.1.3. GHSTS Common Components Schema Definition

In order to avoid inconsistent definition of the same semantics across these XSDs (see above), the auxiliary Common Components XSD contains all XSD type definitions that are reused. The CommonComponents XSD is imported into a these XSDs if a type definition is to be reused.

2.4.1.4. ToC schema definition for ToC templates

The Table of Contents (ToC) used in a GHSTS submission has to have a structure according to a standard Table of Contents defined by authorities (often called ToC templates). The term “Table of Contents” in this document is used as a synonym for “dossier numbering system”.

GHSTS supplies the XSD for the ToC schema definition. This allows authorities or third parties to supply a ToC in a format that it can be used by a GHSTS builder to generate a GHSTS submission.

2.4.1.5. GHSTS Picklist model specification

Upon generation of a GHSTS submission package the registrant has to supply metadata. Some metadata are picklist-controlled, i.e. they must be selected from a list of values where the terminology is pre-defined.

The picklist model specification is a separate document and contains the description of the picklist information model.

Please note: The picklist model specification currently is not developed. The present submission format specification document is not affected by details of the picklist model specification. The interface of this format specification to the picklist model is described in chapter 3.2.24.

2.4.1.6. GHSTS Picklist Model Schema Definition

The GHSTS Picklist Model Schema Definition (GHSTS Picklist Model XSD) contains the technical definition of the GHSTS Picklist information model as an XML representation. This model can be used both by picklist management tools for the management and publication of picklist values as well as for a GHSTS standalone builder tool for the consumption of picklist values as reference data. As both a GHSTS builder and a picklist management tool are considered additional components, this schema definition is not a binding definition for the adoption of GHSTS.

Please note: The picklist model schema definition currently is not developed. The present submission format specification document is not affected by details of the picklist model schema definition. The interface of this format specification to the picklist model is described in chapter 3.2.24.

2.4.1.7. Picklist values (OECD-governed)

Picklist values are to be used in a GHSTS builder environment. The picklists and their values ideally are provided by a picklist management tool (see chapter 2.4.2.3), which should be the primary source for picklist values for a GHSTS builder.

There are three categories of picklists:

- The picklists of type “Core” (GHSTS CORE picklists) are an essential part of the GHSTS specification and are independent of the regulated domain for a submission.
- “Domain” picklists are dependent on the regulated domain for the submission.
- “RA” picklists are dependent on the recipient regulatory authority(ies).

The picklists of type “Core” (GHSTS CORE picklists) and the “Domain” picklists for a submission in the pesticides regulated domain which are not regulatory authority specific, type “GHSTS PPP picklists”, are managed by OECD member countries in the context of the OECD. They are considered GHSTS core components.

The GHSTS CORE picklists and values and a snapshot of the GHSTS PPP picklists and values, made available periodically, will be published on the OECD website (see chapter 7.1).

“RA” picklists and “Domain” picklists and values which are not in the pesticides regulated domain are the responsibility of specific regulatory authority or regulatory domain respectively. They are considered as GHSTS additional components (see chapter 2.4.2.4).

2.4.1.8. ToC templates (OECD-governed)

ToC templates serve to harmonise the Table of Contents (Dossier data points in a hierarchical structure) of a submission by their reuse. They are supplied in the form of XML files following the ToC schema definition (see chapter 5.1) and can be used in the following context:

- Import of the ToC template into a GHSTS builder – to use the ToCs as templates for an upcoming submission. When applied
 - the ToC template XML is copied into the GHSTS backbone XML below the element “ToC”. (Note: The ToC template carries slightly more information than the ToC in the submission package, for example information on the date from when until when the template should be used.)
 - the file of the ToC template is referenced from the GHSTS backbone XML. This serves to allow a GHSTS validator to validate the ToC in the submission against the ToC in the template.
- Usage in authority in-house systems.

The dependencies between ToC template, GHSTS builder and GHSTS submission package are depicted in the following figure.

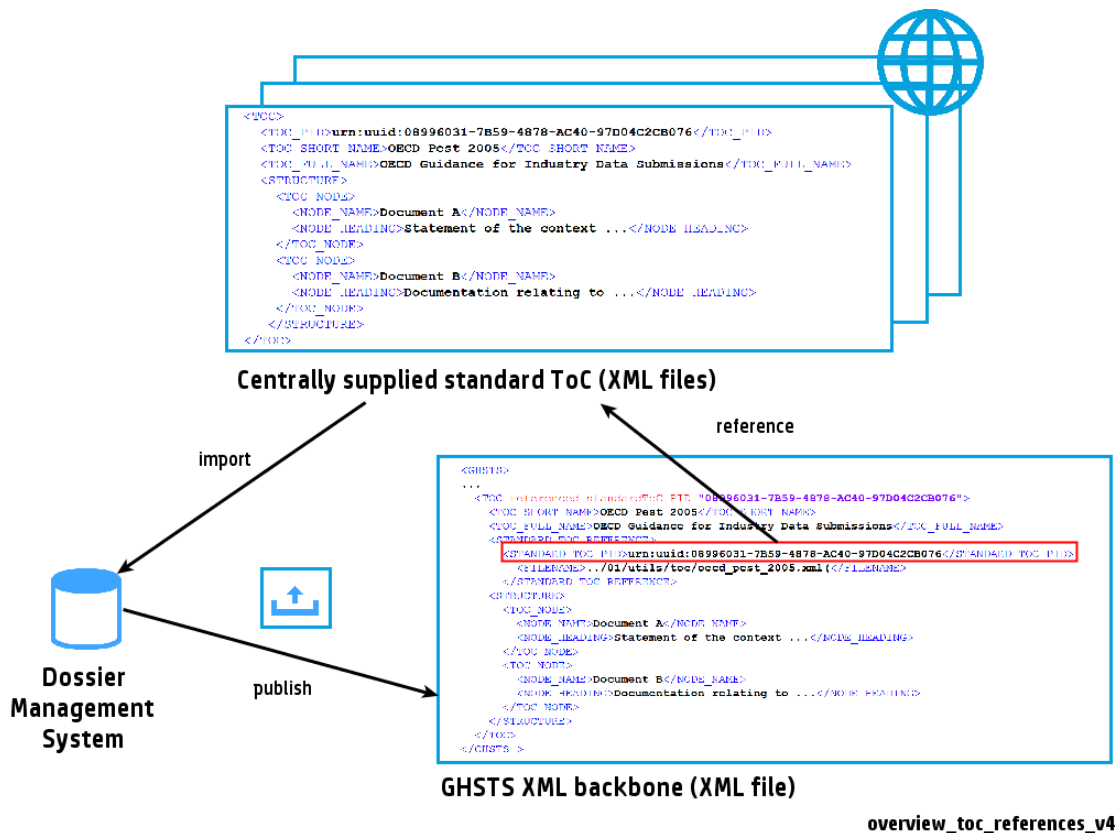


Figure 5 – Overview of ToC references

The top part shows the published ToC template files. These can be imported into a GHSTS builder. During a GHSTS publication process the resulting GHSTS XML backbone file will contain a reference to the used ToC template in the form of a persistent identifier of the ToC. In addition, the ToC template file is also contained in its XML format in the submission package and referenced from the GHSTS XML backbone file (not depicted here).

To support the use case of submissions where no supplied ToC template is available the logical and physical ToC references in the GHSTS XML backbone files are optional. In this case the GHSTS XML backbone will contain a ToC without references to a ToC template file and will not contain an XML file of the ToC template. In this case, no validation between the ToC in the XML backbone file and a standard ToC is supported.

The GHSTS does not provide conversion tables, also called “cross-walk tables”, between different tables of contents. Currently GHSTS focuses on the submission transport aspect with a fixed ToC.

All OECD-defined Table of Contents templates are considered core components, ToC templates for other bodies are considered GHSTS additional components (see chapter 2.4.2.5)

The OECD website will publish information about available ToC templates on the OECD GHSTS website (see chapter 7.1).

2.4.1.9. GHSTS Static Representation XSLT

Every GHSTS submission has to include a static representation of the contents of the GHSTS XML file. This HTML representation is kept deliberately simple (for example, containing no JavaScript and no advanced features) in order to avoid long-term technical issues when displaying the HTML in future versions of web browsers and operating systems.

OECD provides an XSL transformation (XSLT) file that can be used to technically convert the XML backbone file to HTML. A GHSTS builder can directly use this XSLT file, or implement a different method, provided that the resulting HTML file is identical to the one generated with the official XSLT file.

2.4.1.10. GHSTS validation rule specification

A working GHSTS validator technically consists of two components, a GHSTS validator runtime (see chapter 2.4.2.6 – not part of the GHSTS core components) and a ruleset defining the validation rules to be applied. A GHSTS validator runtime applies a selected ruleset to validate a GHSTS submission against the GHSTS specification.

OECD supplies a specification for a technical validation ruleset that ensures that a GHSTS submission package can be technically processed according to this specification. The implementation of this ruleset to be used within a validator runtime is not in scope for OECD.

It is suggested that companies apply a standard technical validation according to this specification prior to submission, as well as authorities upon receipt. Then, authorities may optionally apply additional regulation-specific validations on both metadata and content files, for example, dependent on the application type and regulation. Those specific validations are out of scope for GHSTS as submission transport standard, but should be publically available.

Please note: The GHSTS validation rule specification currently is not developed. The GHSTS standard can be adopted without the availability of the GHSTS validation rules specification. At minimum it is recommended to validate the XML backbone against the GHSTS schema definition, to follow the constraints in the respective subchapters of each XML element and to consider the package definition in chapter Chapter 6.

2.4.2. GHSTS additional components

This chapter describes GHSTS components that potentially support the use of the GHSTS but that are not supported by the OECD member countries. Please refer to the OECD GHSTS website (see chapter 7.1) for current information about available additional components.

2.4.2.1. GHSTS viewer

The term “GHSTS viewer” is used to denote a separate component to view a submission package. It may contain more advanced functionality than the static HTML representation (see chapter 2.4.1.9).

2.4.2.2. GHSTS standalone builder

A GHSTS builder is a software application that can generate GHSTS-compliant submission packages.

2.4.2.3. Picklist management tool

A picklist management tool is a software that can manage picklists and picklist values according to the picklist model specification (see chapter 2.4.1.5). The tool should include the support of:

- lifecycle of picklists
- lifecycle of picklist values
- import of picklist values from other sources
- export of picklist values in technical and human-readable format
- publication of minor and major releases of picklists.

2.4.2.4. Picklist values

“RA” picklists and “Domain” picklists and values which are not in the pesticides regulated domain are not managed by OECD and are in the responsibility of specific regulatory authority or regulatory domain respectively.

2.4.2.5. ToC templates

The ToC templates coming from bodies other than the OECD are considered as GHSTS additional components.

The OECD website will publish information about available ToC templates in the pesticide domain on the OECD GHSTS website (see chapter 7.1).

2.4.2.6. GHSTS validator runtime and rulesets

The GHSTS validator runtime is a software that can validate a GHSTS submission according to defined rulesets.

A GHSTS validation ruleset validates a GHSTS submission package according to specified rules on a defined GHSTS validator runtime. As there may be different GHSTS validator runtimes there may be different formats for the same ruleset that may not be interchangeable. The rulesets may be domain dependant.

OECD defines a technical validation ruleset (see chapter 2.4.1.10), but does not supply an implementation of the ruleset for a specific validator runtime.

2.4.3. Lifecycle and dependencies of GHSTS core components

Each GHSTS core component listed above can have its own lifecycle and will be versioned separately from other components using separate version information. However, the components may still be dependent on each other.

The GHSTS XSD, GHSTS Common Components XSD, GHSTS ToC XSD, and GHSTS Format Specification, are always released together in a “GHSTS release package”. A new GHSTS release package is generated when one of these GHSTS components is versioned. Other components are released with their own versioning as appropriate and according to dependencies with other components.

The dependencies between the GHSTS core components in a Release Package are documented in the following form:

GHSTS core component	Release Package 1 <version and date>	Release Package 2 <version and date>
Core component 1	< version and date>	< version and date>
...	< version and date>	< version and date>
Core component N	< version and date>	< version and date>

Table 2 – GHSTS Release Package Matrix example

One line is maintained for each GHSTS component. Each cell contains the latest version number and date that supports the specific GHSTS release package.

For example, if a new version of the schema definition is created then a new release package is generated. If the change of the schema is irrelevant for a component which is part of the Release Package, for example the GHSTS ToC XSD, then the version of that component will not change compared to the predecessor release package. Comments can be added to the cells, for example, to explicitly specify what min and max GHSTS release package version is supported by a respective component.

The GHSTS release package matrix is available on the OECD website (see chapter 7.1).

Instances of Table of Contents (e.g. OECD ToC, specific national ToC), picklist values, or validation rules will likely change independently of GHSTS release packages and hence be versioned and published separately. There may be dependencies if changes in the GHSTS release package may trigger an update of a component, for example, if a change in the ToC XSD requires that the ToC templates be updated to be compliant with the updated ToC XSD.

The maintenance and documentation of the lifecycle and dependencies of GHSTS additional components is the task of the respective providers of the components.

Chapter 3. Introduction to GHSTS

This chapter introduces GHSTS. It gives an overview for GHSTS mainly from the business perspective, avoiding many references to the technical chapters that follow.

- First in chapter 3.1 the design principles behind GHSTS are outlined.
- Then, the logical concepts that are embodied in the GHSTS submission package are explained in chapter 3.2.

3.1. GHSTS Design Principles

3.1.1. Focus on administrative metadata

GHSTS is a transport standard that can be used across different regulated domains and regulations. The set of metadata that is part of the GHSTS XML backbone is of a generic administrative nature, meaning that the metadata serves to clearly identify the regulatory action and the associated documents. The GHSTS metadata is not meant to transport comprehensive scientific information about the contents of the submitted documents. This is the task of the (structured or non-structured) content (e.g. study summaries, study reports) within the submission.

The technical validation ruleset supplied as a GHSTS core component (see chapter 2.4.1.10) ensures that the GHSTS submission can be technically processed according to the Specification; it does not cover validation on metadata values other than supported by the core components (e.g. validation against CORE picklist values is supported). A registrant should be aware that authorities may apply additional validation on GHSTS metadata.

3.1.2. Support for different regulated domains

The first release of GHSTS in 2014 was designed to support submissions for pesticides. Starting from version 2, GHSTS is opened to be flexible to support the submission process in multiple regulated domains.

The different semantics requirements for different regulated domains are managed by the “Variation with picklists” principle, explained in the following chapter.

3.1.3. Variation with picklists

GHSTS submissions for different regulated domains have the same structure as defined by the XML backbone schema. However, the values of the non-core picklists supplied in the XML elements may differ and the accepted terms from the picklists may be domain specific.

Therefore, most picklists are organized by “regulated domains”. The “regulated domains” are defined in a core picklist. When creating a submission in a specific domain a GHSTS builder will only offer values from the picklist values that are applicable for this regulated domain and the respective recipient(s).

For more information about picklists and their organisation, please see chapter 3.2.24.

3.1.4. *Parallel viewpoints*

A GHSTS submission package supports two different complementary identification view levels, a human-readable level and machine-readable level.

The human-readable level

- is supplied when using a GHSTS viewer or the static HTML representation. The GHSTS XML backbone file stores the coded and decoded value of the picklist values to allow human-readable display without further references to the picklist representation which was used to build the submission.
- is delivered via the mandatory “main” content of documents that is supplied in PDF format.

The machine readable level

- is supplied by the GHSTS XML backbone file supplying additional technical metadata and references to all supplied files. Most instances of the GHSTS concepts are identified by unique program-readable identifiers. This information can be used by ingestion components to identify and ingest the relevant information for the receiver.
- is delivered by additional optional document content, e.g. supplied in XML format, that can be harvested by suitable computer programs but also displayed for human readers when appropriate display resources (e.g. style sheets) are supplied.

It depends on the file format and the available viewer software of the user whether content of a specific format can be considered as human-readable or not: A Word document is not per se human-readable without the appropriate viewer / editing software.

The following figure visualizes the parallel viewpoints with a few sample types of content (“Main”, “Source”, ”Supplement”), with “Source” being the native representation of the “Main” content (e.g. Word file, OHT XML). Please note that - except for “Main” - the types of content are specified within each regulatory domain and supplied by a domain-specific picklist.

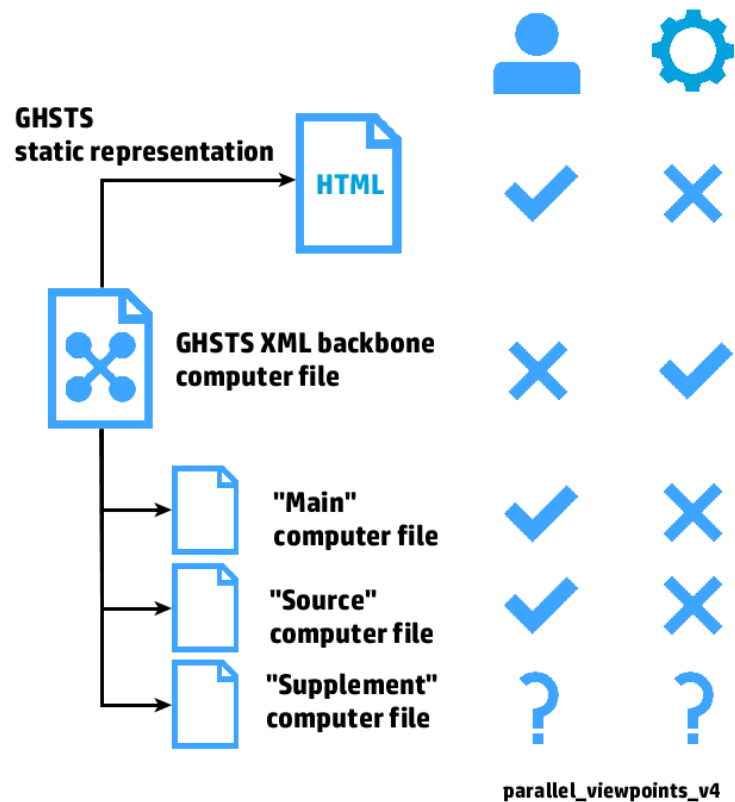


Figure 6 – Support of parallel viewpoints

3.1.5. Parallel navigation scenarios

The metadata in the GHSTS XML backbone supports the two common types of navigation scenarios:

- View by documents/files sorted/filtered indicating the specific regulatory requirements / dossier numbering system (ToC node) with which they are associated.
- View by ToC indicating the associated documents/files.

A GHSTS viewer should support both navigation scenarios in parallel.

3.1.6. Document-based organisation

A GHSTS submission package is organized on two levels:

- A set of DOCUMENTs in which each DOCUMENT incorporates one or more computer files in potentially different formats. The GHSTS concept of DOCUMENT is hence not equivalent to the everyday usage of “document”.
- One GHSTS XML file providing administrative and bibliographic metadata for the DOCUMENTs, the submission, and the guideline / ToC and referencing the set of documents.

3.1.7. Self-contained submission package

The references in the GHSTS XML to DOCUMENTS and FILES must link to electronic files included in the submission package (for full submissions) or the series of submission packages (for incremental submissions) that represent the dossier / regulatory action. It is not permitted to omit document content by including content references in the GHST XML which resolve to URIs outside of the submission

packages of the present dossier, even if specific identifiers (e.g. DOI) could retrieve the document externally. Additional persistent identifiers, for example publicly available content, can be specified in the document metadata and used for citation purposes.

For more details about full and incremental submissions see chapter 3.2.22

3.1.8. Integration of XML-based content

In addition to PDF, DOC, and other content file formats, the GHSTS can also transport XML files containing, for example, structured scientific content like the XML files conformant to the OECD Harmonised Templates for Reporting Chemical Test Summaries (OHT). The information in these XML files can be harvested by computer programs. If the XML document supports standard rendering techniques (e.g. XSL, CSS) and the required assets are part of the submission package and correctly referenced, then the information could also be displayed for users by a GHSTS viewer.

The GHSTS XML backbone contains administrative information. Additional XML files can be transported as attachment files containing scientific study summaries (OHT), labelling information (SMART labels) or other information which needs to be machine readable. However, an accompanying main PDF document would also be required for those additional XML files.

Parties sending or receiving submissions with transported XML files may have to consider potentially conflicting information in the GHSTS XML backbone and the integrated XML content, e.g. for metadata and lifecycle information, or the consequences of (external) references with respect to the self-containment of the submission package.

3.1.9. Table of contents as main structuring element for documents

A table of Contents (ToC) in the context of GHSTS is a hierarchical structure of ToC nodes. Usually the leaf nodes on the lowest level represent specific regulatory requirements for which information has to be supplied by the registrant. Inner nodes serve as “navigation nodes”, e.g. for hierarchical display. Dependent on authorities the depth of ToC structure and the display vary greatly.

The ToC metaphor, originating from paper organisation, has survived in the electronic era and is the dominant way of organising information for pesticide submissions. In GHSTS a document can be assigned to one or many nodes to indicate that this document fulfils the regulatory requirements that are represented by the ToC node and explained in more detail in the underlying regulation.

Each governing body of a regulated domain must take measures to avoid contradictions between the semantics of the ToC nodes and the semantics of picklist-controlled metadata (e.g. DOCUMENT_TYPE) of documents assigned to the nodes.

3.1.10. Adaptiveness with respect to different standard ToC templates

The GHSTS standard is independent of the structure of the Table of Contents. The GHSTS XML backbone can contain an arbitrary Table of Contents (ToC) structure. Optionally a GHSTS validator can check whether the used Table of Contents in the submission package corresponds to one of the existing standard ToC templates.

3.1.11. Lifecycle management

The GHSTS standard contains lifecycle information for different embodied information concepts (for example, product/component, document metadata, document content, and ToC assignments) to allow the quick identification of changes between subsequent submissions of this dossier. This information can be used during review or ingestion to focus on the changes in the latest version.

For details about the content lifecycle management see chapter 3.2.17; for the metadata lifecycle management see chapter 3.2.3.3, for information about the node assignment status see chapter 3.2.20.

3.1.12. Persistent identifiers of information concepts

GHSTS supports the use of unique and persistent identifiers of a set of information concepts that are considered to be valid and reusable outside of the scope of one regulatory action, one registrant or one regulatory authority:

- Document
- Document Family
- File
- Dossier
- Product
- Component (optional)
- Legal Entity
- Standard ToC (optional for ToC nodes not part of a standard ToC template)
- ToC node (optional for ToC nodes not part of a standard ToC template)

These identifiers are called Persistent Identifiers (PID). The term “persistent” underlines that the instances of these information concepts can permanently be identified with this identifier throughout their lifetime and in all contexts where this instance is used. This minimizes the need to send duplicate information by referencing its unique identifier in subsequent submissions.

For most of the above concepts the registrant generates a PID once at the creation of the instance of the information concept and will use this across all regulatory actions. The correct management of the PIDs is under the responsibility of the governing stakeholder of the respective information concept.

If an instance of one of the above information concepts uses a specific PID that has been used and assigned before, then the consumer of the GHSTS submission (human reader or program) can safely assume that the instance is identical to the one submitted before without the need to check the complete concept in detail.

For example, if a document is submitted with a specific PID as part of a submission package, the receiving regulatory agency can verify if it or another agency has already received and evaluated this document before by simply checking the PID against their dossier management system.

For more information about persistent identifier see chapter 3.2.26.

3.1.13. Information specific for regulatory authorities

GHSTS provides the possibility to supply individual and different metadata to regulatory authorities within the same regulatory action when multiple regulatory authorities are involved. For details see chapter 3.2.28.

3.2. GHSTS logical concepts

This chapter explains the different information concepts embodied in GHSTS on a conceptual and semantic level. The technical reference on how these concepts are represented in the submission package, mainly in the GHSTS XML backbone, is contained in the subsequent chapters.

3.2.1. Logical concepts and XML syntax

Although this chapter is about the conceptual introduction of logical GHSTS concepts, some technical XML terms have to be introduced for better understanding.

An element in the GHSTS XML backbone file is used to represent concepts. The following incomplete XML excerpt shows an example:

```
<DOCUMENT id="T-423584-01-1">
  <DOCUMENT_GENERIC>
    <DOCUMENT_PID>urn:ghsts:74343233-3538-3430-3131-000000000000</DOCUMENT_PID>
    <DOCUMENT_TITLE>C 3470 techn. - Acute oral LD50 in the Chinese hamster</DOCUMENT_TITLE>
    <DOCUMENT_AUTHOR>Sarasin, G.</DOCUMENT_AUTHOR>
  </DOCUMENT_GENERIC>
</DOCUMENT>
<DOCUMENT id="T-423586-01-1">
  <DOCUMENT_GENERIC>
    <DOCUMENT_PID>urn:ghsts:74343233-3538-3630-3131-000000000000</DOCUMENT_PID>
    <DOCUMENT_TITLE>C 1983 techn. - Acute oral LD50 in the mouse</DOCUMENT_TITLE>
    <DOCUMENT_AUTHOR>Kobel, W.</DOCUMENT_AUTHOR>
  </DOCUMENT_GENERIC>
</DOCUMENT>
```

Figure 7 – Example of the metadata in XML linked to a document (excerpt only)

An element consists of a start tag (e.g. <DOCUMENT>) and an end tag (e.g. </DOCUMENT>). The element <DOCUMENT> has “element content”, because it can only contain sub-elements as content. The element <DOCUMENT_TITLE > has “text content”.

The element < DOCUMENT_GENERIC > is a *direct* child (or subnode) of the element <DOCUMENT>. The element <DOCUMENT_TITLE> is an *indirect* child (or subnode) of the element <DOCUMENT>, because it is also nested within the <DOCUMENT_GENERIC> element. Elements that are cross-referenced do not count as direct or indirect children.

An element may also possess attributes that appear in the start tag of the element. The element document in the example above possesses an attribute “id”.

Note: The term “attribute” within this document is used only to designate a specific syntax within an XML file. It is not used as synonym for metadata.

For further details about XML please see e.g. https://www.w3schools.com/xml/xml_what_is.asp.

3.2.2. Overview

The following figure shows the concepts of GHSTS - as represented in the GHSTS XML backbone - and the internal relationships that are not specific for an individual receiver. Internal relationships are relationships between elements of the same submission, this means within the GHSTS XML backbone file.

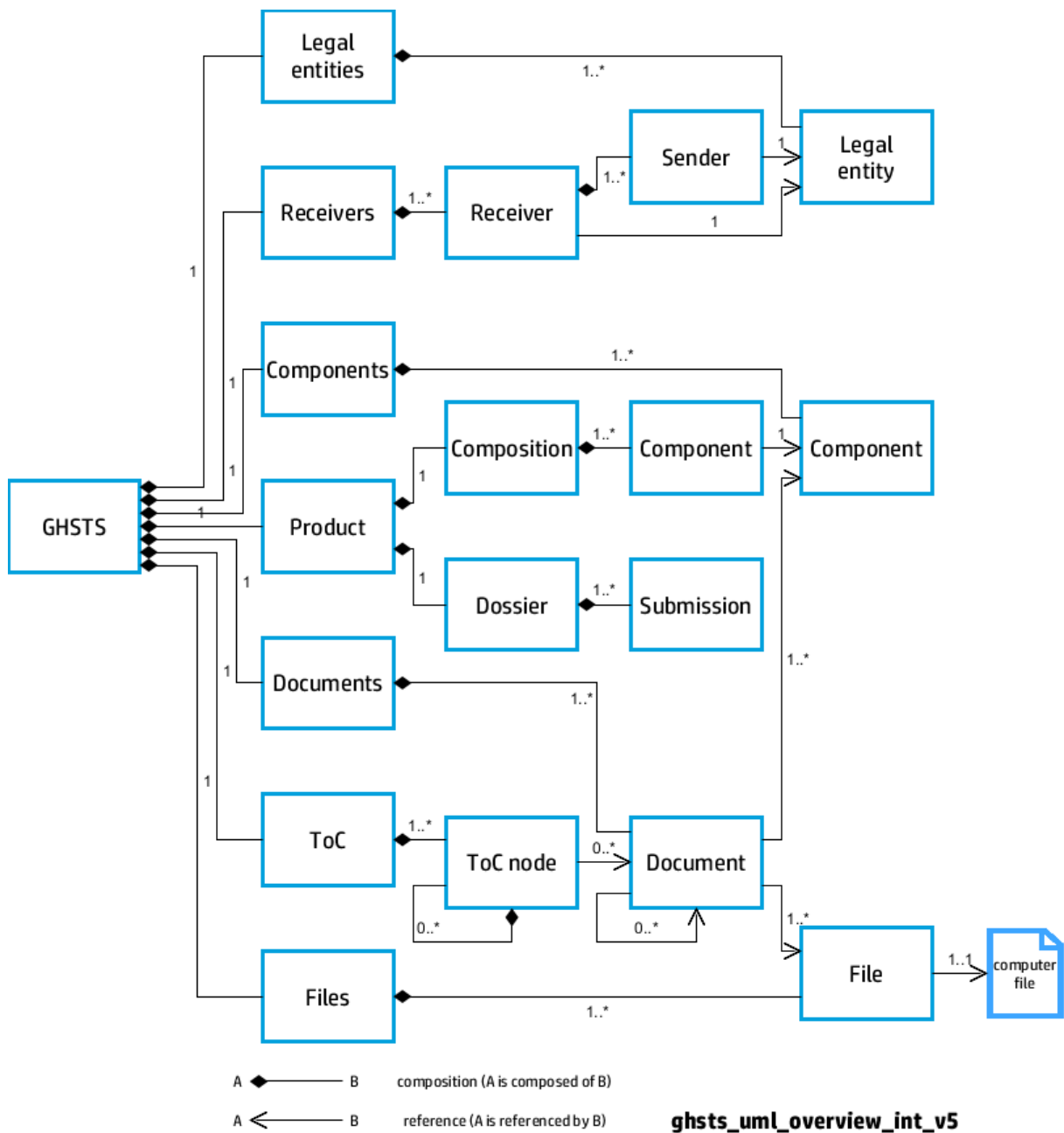


Figure 8 –Main GHSTS concepts and their relations

The figure shows two types of syntactic relations between the XML representations of the main GHSTS concepts:

- The relation “composition” (not the concept “composition” of a product) is represented by nesting in the XML backbone file: if A is composed of B then B is syntactically nested in A in the XML file. E.g. a ToC is composed of ToC nodes.
- The relation “reference” is represented by elements or attributes containing the ID of the referenced element.

For an overview about receiver-specific references see chapter 3.2.28, for an overview of external references see chapter 3.2.28.

3.2.3. Metadata

The term “metadata” or more precise “descriptive metadata” is used to denote information about a logical concept, for example a dossier (dossier name), a document (document title) and other concepts of GHSTS. It is set in contrast to the “content” of a concept (see concepts FILE and DOCUMENT, chapter 3.2.9 and 3.2.12).

For XML-based content the notion of metadata and content is blurred, as both the GHSTS XML backbone containing the metadata and the XML content have the same XML format. In general, the metadata for FILE and DOCUMENT is used to select, identify, cite and retrieve the correct computer file and is mostly a subset of the identifying information that is part of the content of the computer file itself, independent of its file format.

In a submission package the metadata is represented as elements or attributes within the GHSTS XML backbone.

3.2.3.1. Formatting of metadata

Formatting of metadata is not permitted. Special characters like tabs, LF, CRLF, NL and multiple white spaces are not allowed. The same applies to supplying XML fragments as text content of GHSTS elements. The display formatting of the information is up to a GHSTS viewer, the registrant should only supply the unformatted information.

The only exception to this rule is the element LEGALENTITY_NAME where formatting is allowed.

3.2.3.2. Legal metadata characters

The character encoding that can be used to submit information in the GHSTS XML backbone file is UTF-8. With this encoding all characters from the Unicode character set can be represented. The latest version of Unicode contains a repertoire of more than 110,000 characters covering 100 scripts.

All common web browsers are capable of displaying Unicode characters if the required fonts are installed on the computer. Eventually the missing fonts have to be downloaded and installed. Characters for missing fonts are mostly displayed by the sign “□”.

Please note that Unicode includes a variety of options (e.g. right-to-left languages, Chinese, Japanese, and Korean [CJK] fonts) that require the support not only of the XML backbone file, but also of other GHSTS components (GHSTS validator, GHSTS viewer, publication and ingestion component, dossier management system). Therefore it is recommended that the registrant and receiver agree on the features of Unicode that can be used.

As in all XML documents the special characters < (<), & (&), > (>), " ("), and ' (') need to be encoded in the GHSTS XML backbone when used for element or attribute values.

3.2.3.3. Metadata changes

GHSTS supports the tracking of metadata changes for a number of concepts that are represented in the GHSTS XML backbone. The following values are distinguished:

Value	Description
New	The metadata belongs to a new instance of the concept that has not been part of the previous submission. In the first submission 'New' should be used in all metadata status elements
No change	The metadata of the instance of the concept did not change compared to the previous submission.
Modified	The metadata of the instance of the concept changed compared to the previous submission; the instance itself is still part of the submission.

Table 3 – Values for metadata changes

The tracking is done by setting one of the above values to the element METADATA_STATUS of the concept. Usually the metadata status is set for all elements in the GHSTS XML backbone that are within the same block or its children. However there are some exceptions to this rule. The following table lists the concepts within GHSTS that possess the element METADATA_STATUS and explains the scope of this element.

Concept	Description of concept	Scope of tracked metadata changes
LEGAL_ENTITY	Complete Information for receiver and sender see chapter 3.2.8	Any element within LEGAL_ENTITY
RECEIVER	Specific additional metadata for receiver and sender, added to information in LEGAL_ENTITY	Any element within RECEIVER, including the child SENDER concept.
COMPONENT	Concept for component	Any element within COMPONENT
PRODUCT	Concept for product, referencing the component concept	Any element within PRODUCT and all direct and indirect children, except for the SUBMISSION element. This includes changes to the set of used ingredients and their quantity. The SUBMISSION element changes by default for every submission; hence these changes are not reported.
FILE_RA	Concept for RA-specific metadata on the FILE level	Any element within FILE_RA
FILE_GENERIC	Concept for generic metadata on the FILE level	Any element within FILE_GENERIC except for the block CONTENT_STATUS that contains the information for content changes.
DOCUMENT_RA	Concept for RA-specific metadata on the DOCUMENT level see chapter 3.2.12	Any element within DOCUMENT_RA
DOCUMENT_GENERIC	Concept for generic and bibliographic metadata on the DOCUMENT level see chapter 3.2.12	Any element within DOCUMENT_GENERIC except for a) the block DOCUMENT_CONTENT_STATUS_HISTORY that contains the information for content changes b) references to the FILE concept (REFERENCED_TO_FILE) that can change with every (full) submission. The scope includes changes to references to other documents or references to components

Concept	Description of concept	Scope of tracked metadata changes
TOC	see chapter 3.2.7	Any element on the same nesting level as the METADATA_STATUS element and their sub-trees (e.g. TOC_SHORT_NAME, TOC_OWNER, TOC_FULL_NAME, TOC_VERSION, STANDARD_TOC_REFERENCE), except STRUCTURE. Changes in and below STRUCTURE are not tracked by METADATA_STATUS.

Table 4 – Concepts that track metadata changes

3.2.4. Dossier regulatory action

A dossier regulatory action is a series of events accomplished by a regulatory authority beginning with the submission of an application with data by a registrant and ending with a unique regulatory decision. During one regulatory action the applied standardised Table of Contents (for example, OECD, EU, national) has to remain the same. It is up to the communicating parties to decide what kind change of

- Registrant
- Regulatory authority (RA)
- Component
- Product

could be covered under the same regulatory action or requires a new regulatory action.

Within GHSTS a dossier regulatory action is identified with a PID called DOSSIER_PID which remains unchanged throughout all submissions. In addition, GHSTS supports the identification of a dossier with an identification assigned by the regulatory authority using the element PROJECT_ID_NUMBER.

3.2.5. Submission

A submission is the compilation of documents in a structured form according to the given regulatory requirements, as part of a dossier regulatory action.

The physical representation of a submission on media is called submission package.

In GHSTS a submission within a dossier regulatory action is identified with the SUBMISSION_NUMBER.

The concept of submission in the GHSTS XML backbone file is represented by the set of metadata, consisting of submission number, submission version date, submission title and the information whether the submission is incremental or not (see chapter 3.2.20).

Each submission package contains this information for the current and all preceding submissions, if any, of the regulatory action.

3.2.6. Dossier

A dossier is the set of information collected by the registrant for the purpose to serve a specific regulatory action and updated according to the process of the regulatory action. A GHSTS submission captures the state of a dossier at a certain milestone in GHSTS format.

For each regulatory action the registrant has to create a separate dossier with a new DOSSIER_PID.

3.2.7. Table of Contents

Table of Contents (ToC) are defined by regulatory authorities / regulatory framework and order the documents within the submission. Within a specific regulated domain there may be multiple ToCs for different regulations and application types. A specific ToC instance may also be used across regulated domains.

3.2.7.1. ToC versioning

A ToC may be versioned over time, resulting in a ToC family and multiple versions of the same ToC. A ToC carries the same TOC_SHORT_NAME across all versions and a specific TOC_VERSION_TITLE and version string for each version. Each ToC version has its own PID to support machine readability and submission processing.

Each ToC version within a ToC family has to fulfil the “grow constraint”. This means that ToC nodes that were present in a previous ToC version must not be changed in the subsequent version, and ToC nodes may only be added to a new ToC version. Otherwise, it would not been possible to use a later ToC version in a subsequent version of the same regulatory action, as documents that had have been assigned to nodes being deleted or changed would be orphaned.

If the regulatory requirements require the deletion or change of ToC nodes then this would be modelled as version 01.00.00 of a new ToC version family.

The Table of Contents (ToC) that is used as part of the GHST XML backbone is identified by the element TOC_SHORT_NAME.

The registrant has to use the appropriate ToC from the predefined ToC family, usually the latest published version. Upon agreement of sender and recipient, it is permitted to use a subsequent ToC version in a later submission of the same regulatory action, but it is not permitted to change the ToC family within the regulatory action, due to the potential of creating “orphaned documents” described above. A ToC template version can optionally carry a date TO_USE_UNTIL, if a temporary nature of the ToC template is clear from the beginning.

Each GHSTS backbone XML file contains a complete copy of the standard Table of Contents, even if not all ToC nodes are filled. This requirement is the pre-condition to process or display an old dossier years later with exactly the same context, which was valid at the time of submission. That means “freezing” the full ToC into each GHSTS submission package makes GHSTS non-sensitive regarding the version management of standard TOC.

The GHSTS backbone can contain an optional reference to the XML file of the applied standard ToC. This reference can be used by a validator to check whether the embodied ToC in the GHSTS backbone XML is identical to the applied standard ToC. The reference is represented by a logical PID, by the TOC_SHORT_NAME and the physical reference to a ToC XML file (see chapter 4.9.37) that has to be part of the submission package.

3.2.7.2. OECD tables of Contents

The OECD encourages regulatory authorities to adopt the use of the OECD Table of Contents for dossiers in order to allow for easier communication among regulatory authorities during Joint Reviews and to reduce the regulatory burden on applicants. The OECD Table of Contents can be utilized for both joint and many national submissions.

The OECD makes available different dossier ToC in XML format according to the ToC XSD described in chapter 5.1 in order to be incorporated easily into the meta-data of an electronic submission package

built according to GHSTS. For the list of available ToC and where they can be found please see chapter 2.4.2.5.

3.2.7.3. National / Regional tables of Contents

Though the OECD encourages regulatory authorities to adopt the use of the OECD numbering system for dossiers, other numbering systems exist and national/regional applications of different types (for example, end-use product authorisations with active substances already evaluated, for use extensions of existing products, for MRL setting, for EU zonal authorisations or mutual recognition) remain a major workload for most OECD member countries. GHSTS serves as a submission tool for these other types of applications and can incorporate national numbering systems or the OECD numbering system in the meta-data of an electronic submission package to describe the files being transported.

To this end, Regulatory authorities have the responsibility to make their national/regional ToC templates available in XML format to the OECD, if they choose to have it available to be incorporated easily into the meta-data of an electronic submission package built according to the GHSTS. For the list of available ToC and where they can be found please see chapter 2.4.2.5.

It is recommended that, if a dossier is submitted to more than one regulatory agency, the appropriate OECD ToC (for chemicals, micro-organisms or pheromones) should be used rather than using a national ToC. Ultimately, regulatory authorities decide which ToC should be used.

3.2.8. Legal entity

The concept of legal entity summarizes all contact information both for the registrant(s) and regulatory authorities.

Whether an entry for a legal entity serves as information for sender or receiver is specified separately in the concepts sender and receiver by cross-referencing the corresponding entry. Sender and receiver only contain few additional metadata to specify their role.

GHSTS supports the definition of multiple receiver and for each receiver the definition of individual multiple sender.

3.2.9. Product

A regulatory action serves to receive a regulatory decision for a certain product. In the pesticides domain there are two kinds of products:

- A product as a formulated product / composition of multiple ingredients. An ingredient in the GHSTS backbone XML is the reference to a component (see chapter 3.2.10) specifying its quantity in the product.
- A product as Technical Grade Active Ingredient (U.S.)/Active Substance (EU). In this case the product name is set to the common name of the Active Ingredient / Active Substance and the only ingredient of the product is this active substance.

These two kinds of products are reflected by two types of dossiers with different studies:

- Dossiers for active ingredients, for example Technical Grade Active Ingredient (U.S.)/Active Substance (EU)
- Dossiers for formulated products

In other regulated domains, the term “product” designates the object that is subject to regulation in the specific domain, for example, this can be a product, substance, claim, organism or process which requires a dossier submission as part of an application for authorisation.

In the pesticides domain the following rules apply when using the product concept: Each product or TGAI (Technical Grade Active Ingredient) is identified by a PID. For example, the TGAI “saflufenacil” has a PID and has the generic product name “saflufenacil” in the GHSTS submission. In addition the registrant can supply specific product names for each receiving regulatory authority, for example.

- BAS 800 H,
- Saflufenacil,
- Saflufénacil,
- saflufenacil TC,
- benmihuangcaoan.

Within the GHSTS XML backbone the product is identified with a unique identifier that will be created by the submitting company. This identifier cannot be changed during the ownership of the company; hence the identifier has to be reused for all regulatory actions concerning this product. The identifier of the product designates the ingredients in the product and is independent of the product name, so the unique identifier will remain stable even when the product name changes. A new product identifier has to be generated when the composition of its components is changed.

3.2.10. Component

In the GHSTS XML backbone for a submission the registrant has to list all relevant components that are part of the product.

Components can be substances and/or organisms (see chapter 3.2.24.1) depending on the regulated domain. For example, in the pesticides domain, these are the chemical substances as active ingredients as well as the metabolites and impurities of the product. Impurities of active ingredients and co-formulants do not have to be supplied as part of the GHSTS XML backbone.

The components are referenced by the product in the GHSTS XML backbone. In the pesticides domain this reference can be considered as ingredients of the product. Components can also be optionally referenced by single documents.

Components can be divided into different component classes, controlled by a cross-domain core picklist (see chapter 3.2.24.1).

Each component is identified / described by two types of information. Each type of information can be supplied multiple times for each component:

- The component identifier allows a unique identifier from an international register or source to be linked with the component. Type is picklist controlled list of recognised sources of identifiers for e.g. Chemical Abstracts Service or American Type Culture Collection. Value contains the unique number of code.
- The component description allows taxonomic or other characteristics of the component to be reported. Type is picklist controlled list of descriptors for example, "Species". Value contains the description in text, for example "Zea mays".

The component is also (optionally) identified with a unique PID identifier (see chapter 3.2.26). In addition, the registrant supplies a common name of the component if available, otherwise an internal name.

3.2.11. File

In the context of an electronic submission the term “file” is usually used as synonym for computer file, as being the basic resource for storing information on a computer. Each computer file possesses a content plus some additional file attributes (e.g. size, file creation date). The latter are managed by the operating system.

For GHSTS this concept of “file” is not sufficient. It is not possible to easily model additional GHSTS-specific metadata and to track the lifecycle of a computer file during a submission. Therefore, GHSTS contains the concept of a FILE (in capital letters).

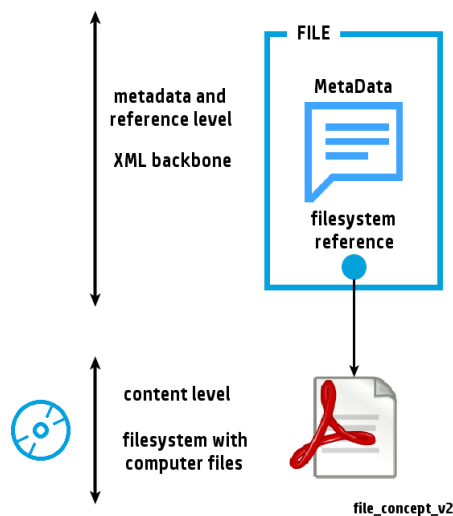


Figure 9 – logical concept of FILE

As shown in the preceding figure, a FILE is a concept within the XML backbone and is composed of

- Metadata (e.g. a file ID supplied by the company)
- One single reference from the XML backbone to one file on the file system (a PDF file in the example)

In the remainder of this document, the term “FILE” will be used to identify this GHSTS concept, whereas “file” will be used as synonym for computer file.

3.2.12. File type

A FILE never appears on its own but always as part of its parent compound concept called DOCUMENT. A DOCUMENT instance can have one to many FILE instances.

The role of each FILE in its parent DOCUMENT is defined by the value of the element FILE_TYPE. The value of this element is different and independent of the technical file format (PDF, XML), but is specific to the meaning and purpose of this FILE in the context of its parent document.

The possible values of FILE_TYPE are defined by the regulated domain and are therefore defined in a non-core picklist. The following table provides suggested values for FILE_TYPE which are divided into two classifications: one core representation and different types of attachments. Additional values for FILE_TYPE can be used for a specific domain if required.

Type of file	Classification	Interpretation
Main	core representation for the regulatory review, containing the full content	The FILE instance references a computer file in PDF/A format. It contains the full content information. This is the primary document for the regulatory reviewer.
Source	Attachment An attachment can contain either full, partial or supplemental content information compared to the content of the core representation.	The FILE instance references a computer file containing the same full content and layout as the “Main” computer file in its native and editable format, e.g. in Microsoft Word format. “Source” and “Main” content in one DOCUMENT instance will always be changed simultaneously. An instance of an OECD Harmonised Template (OHT) in its native XML format would be supplied with FILE_TYPE “Source”, and a PDF rendition of the OHT would be supplied as “Main”.
Supplemental		The FILE instance references a computer file with supplemental information for the review, e.g. the raw data for an evaluation, high resolution pictures, data files for statistic programs, Excel-Sheets, etc. The information in the Supplemental will never change independently of the Main or Source file type.

Table 5 – Overview of types of FILE

Please note the following constraints:

- The value “Main” has to be present in each regulated domain-specific picklist for FILE_TYPE
- There must be exactly one FILE instance within each DOCUMENT instance with FILE_TYPE = “Main”. This means that all RA to which the DOCUMENT is assigned will get the same core representation. If different RA need to have different variations of the same core content (e.g. by language or formatting), create different DOCUMENT instances with different PID, make them part of the same DOCUMENT_FAMILY (see chapter 3.2.13) and optionally link the documents using the element REFERENCED_DOCUMENT.
- Every DOCUMENT instance must be contained only once in the submission. Each DOCUMENT instance must be linked to one or many ToC nodes (data points).
- If it is necessary to split one "Main" document into different files to minimize the file sizes (for example of large toxicological study reports) this can be managed by assigning the first file as “Main” and the subsequent parts as “Supplemental”. If files other than "Main" are split, then they can carry the same value for FILE_TYPE.

3.2.13. Document

An instance of a GHSTS DOCUMENT is a compound object of one to many content representations, modelled in GHSTS in the FILE concept plus the metadata of the DOCUMENT itself. A sample instance of a DOCUMENT is shown in the figure below:

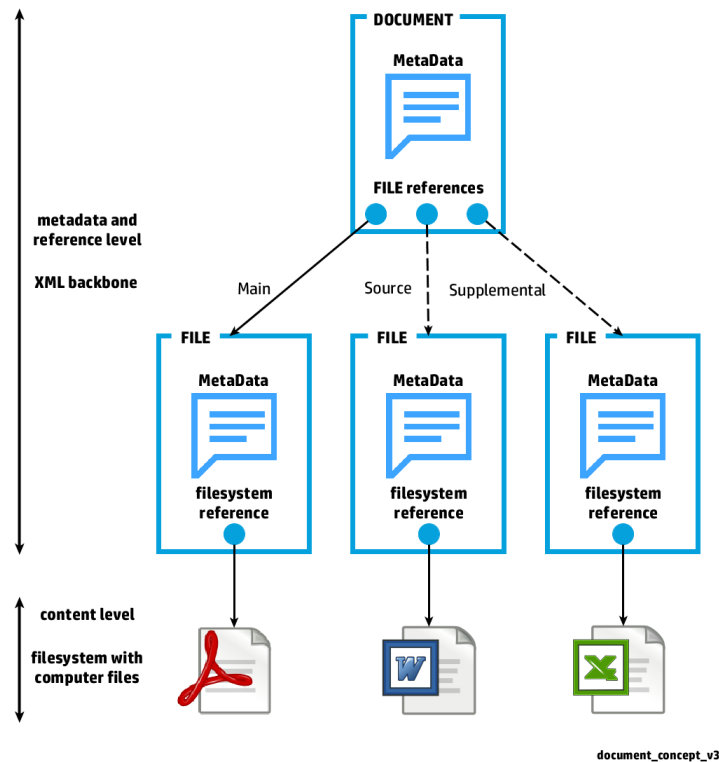


Figure 10 – logical concept of DOCUMENT

The “Main” representation of the DOCUMENT is the core content element. Its content and content classification determine the metadata of a DOCUMENT which is inherited to all its attached FILE instances (e.g. the DOCUMENT_TYPE, see chapter 3.2.15). The other content elements of classification “Attachment” are additional and auxiliary content elements to support the content of the “Main” representation such as a study report in PDF format which can be grouped with:

- its source representation in Microsoft Word
- high-definition renditions of pictures from the study report
- Excel files with data present in the study report

The following principles determine what content to group within a DOCUMENT instance:

- All “Attachments” have a content lifecycle dependent on the “Main” representation
- All “Attachments” complement and detail the content of the “Main” representation.
- An “Attachment” will not change its content when the “Main” content is not changed. (This is true with the exception of apparent error corrections, for example when wrong attachments are replaced with correct ones).

From an information viewpoint the information of each FILE attachment is directly dependent on the information in the Main file. The scientific evaluation of a DOCUMENT instance is not dependent on changes in its “Attachment” parts. This means that either the attachment contains virtual the same information as the (PDF) Main file in a different technical format (e.g. Word or XML) or that the attachment contains information which is never required to be referenced outside of the context of the

Main file of this DOCUMENT instance (e.g. a chromatography picture extracted from the Main file) and will never change independently of the Main file.

This principle defines the possibilities and limits of the grouping of content within one DOCUMENT instance

- As DOCUMENTs are assigned to ToC nodes and not FILEs, the grouping feature of a document avoids the proliferation of content entries within a ToC.
- FILE Content with independent lifecycle cannot be grouped in the same DOCUMENT. For Example, a study report and a study summary (in whatever formats) cannot be part of the same DOCUMENT instance, as a study summary may change independently of the study report due to new scientific requirements and results, and also vice versa. In addition, the DOCUMENT_TYPE classification on DOCUMENT level prohibits a mixing of multiple FILE instances of different document types in one DOCUMENT instance.
- Translations of the same information (e.g. a specific study report) are to be put into individual DOCUMENT instances, as
- internal and external identifier are attached to the DOCUMENT level (e.g. DOI, RA document number) that vary across translations
- in joint-review scenarios different translations can only be targeted to different regulatory authorities on the DOCUMENT level (see chapter 3.2.28), not on the FILE level

The following rules apply for the complete content lifecycle of a specific FILE instance with a specific PID, independent of regulatory actions or submissions:

- There can be only one ‘Main’ file in a DOCUMENT instance. A FILE instance cannot be categorised as “Main” in one DOCUMENT instance and “Attachment” in another, as otherwise the role of "Main" as defining content for the DOCUMENT would become ambiguous.
- If a specific FILE is used exclusively as “Attachment”, it can be part of multiple DOCUMENT instances. For example, the same FILE instance may be part of two DOCUMENT instances as “Attachment”, one DOCUMENT instance being the sanitized and the other being the non-sanitized variation of the same study report. When both such DOCUMENT instances are used within the same submission, one FILE instance may hence be referenced by multiple DOCUMENT instances also in the GHSTS XML backbone file.

In order to group content that is related, but modelled into different DOCUMENT instances, the respective instances can be cross-referenced using the element REFERENCE_TYPE between DOCUMENTS with an appropriate picklist value. Related DOCUMENTS of document type study report and study summary can be linked this way, and a reviewer can be guided from a study report to the related study summary / summaries independently of their assignment to different ToC nodes.

The composition of the attachments of a DOCUMENT instance may change over time, but the information in the attachments is only associated with the Main file. There may be attachments added or replaced over time even when the Main file stays constant, but even then those attachments are bound to the information in the Main file only, and are not relevant for the reviewing process as isolated information concepts.

It is advisable that specific regulated domains define additional specific rules for further information concepts on how to group FILE instances to DOCUMENT instances and to cross-reference DOCUMENT instances. This definition is outside of the scope of GHSTS.

3.2.14. Document Family

A semantic content change of the core representation of a DOCUMENT triggers the creation of a new DOCUMENT instance with a new PID. There are two types of content changes:

- Evolution in content, for example when a study report is amended or content of a study summary is updated due to latest scientific insights. This can also be called “versioning of content”.
- Variation in representation, for example when a study report is translated into a different language or a new layout of a study summary needs to be created (e.g. different header pages, stamping, signature policies).

Yet still all changes are variations of the same (ontological) information. A study report about acute oral toxicity will never be updated with a study summary of physical properties.

GHSTS possesses the concept of a Document Family where all DOCUMENT instances resulting from those changes are grouped. The following figure shows the concept.

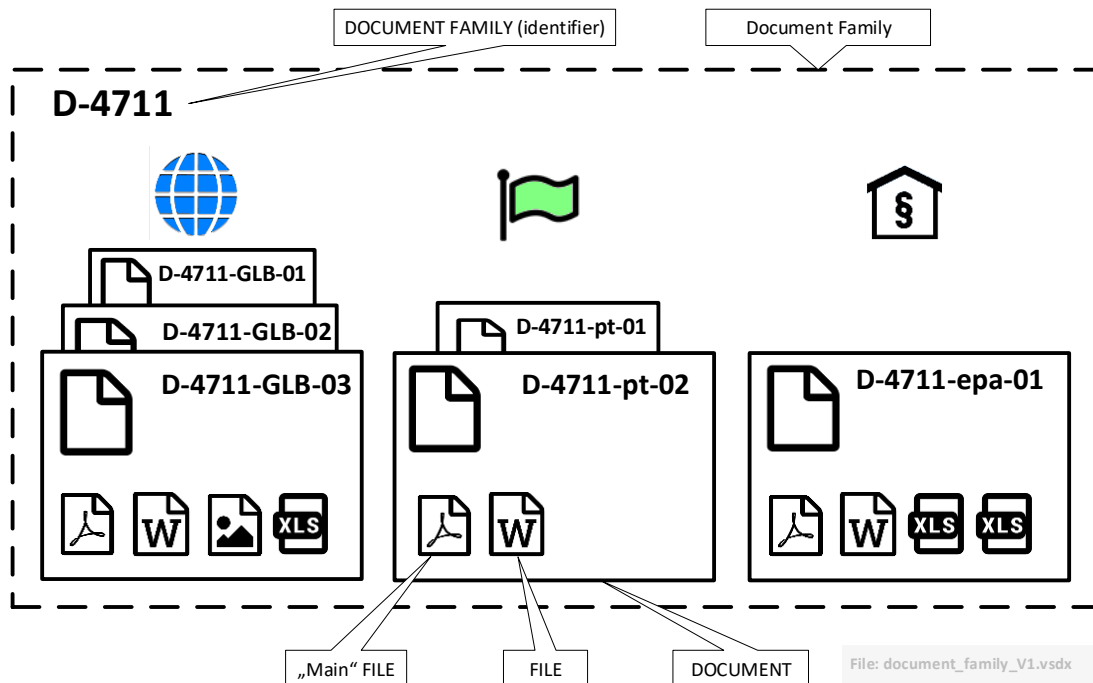


Figure 11 – Document Family, Document and File

The figure shows three sample variations of a specific information, for example a specific study report, representing one document family: A global study report layout (GLB), a Portuguese translation (pt) and an EPA specific layout variation (epa). The content of the respective FILE instances and the composition of the DOCUMENT follow the respective regulatory requirements. The documents for each variation can be lifecycled independently over time (-01, -02, -03). A document family can be considered a 2-dimensional version matrix compared to the linear version chain in a standard document management system. A DOCUMENT instance becomes unalterable once a later version exists or it has been submitted to an authority, as its PID, as unique identifier, represents a specific content and needs to remain stable over its lifetime.

GHSTS technically represents the concept of a document family with two depending elements as synonyms:

- The element DOCUMENT_FAMILY carries a short textual identifier
- The element DOCUMENT_FAMILY_PID carries a unique identifier for the document family.

For each DOCUMENT instance it is recommended to use the value of the element DOCUMENT_FAMILY as leading string of the value for the element DOCUMENT_NUMBER.

Both elements carry the same additional information that can be used during viewing, filtering or machine processing.

The Document Family concept is relevant for the “replace” operation of a DOCUMENT in a ToC node of a dossier. A DOCUMENT instance can only be replaced in the ToC node by another instance of the same document family (see chapter 3.2.18.). If a DOCUMENT instance is to be exchanged with an instance of a different document family, this has to be represented as deletion of the first instance and adding of the second.

The values of both identifiers DOCUMENT_FAMILY_PID and DOCUMENT_FAMILY act as unique lifetime identification for a document family. Each document family needs to possess a unique and distinct pair of such identifiers.

As documents, the document family and PIDs are mandatory basic concepts in GHSTS it is required to apply corresponding concepts and policies in the document management systems of the registrant as well as the receiving authority.

3.2.15. Document type

Each DOCUMENT instance can optionally carry a DOCUMENT_TYPE information, controlled by a non-core picklist (see chapter 3.2.24.2).

- During viewing this information can be used to quickly identify document categories and to filter / subdivide document instances in a ToC node, if needed .
- During validation it can be used in business-level validation processes to ensure a submission is complete.
- During processing it can be used by the receiver to process and store the received files.

Please note that the DOCUMENT_TYPE information is part of the DOCUMENT metadata, but it is semantically linked to the document family concept. Therefore, its value must not differ between the documents within the same document family.

3.2.16. File formats

The term file format describes the encoding of information in a computer file. GHSTS is not an archiving standard, but a transport standard, however the aspect of long-term usability of the content in the computer files should be considered.

GHSTS recommends supplying the “Main” file format as a specific PDF format called GHSTS PDF profile. The characteristics of this PDF format are described in a GHSTS PDF profile in the next subchapter.

For the computer files of other file formats there is no imposed constraint. If possible GHSTS recommends using file format types that are established, standardized, and widespread, and support a long-term visualization and lossless migration. It is expected that word processing, spreadsheet and XML files will be used.

The following characteristics of file formats are not recommended:

- A computer files that is zipped
- A computer file representing an email message (e.g. an Outlook .msg file)
- A password protected, encrypted or digitally signed computer file

Regional business rules can add or alter those constraints.

If a user opens the computer file on a computer where the file extension is not known usually a message box will pop up and the user is requested to select an appropriate viewer application. Therefore the registrant should

- only use file formats where he is sure that the receiver has the appropriate viewer / editor software already installed,
- supply information about the file format in the element MEDIA_TYPE. (see chapter 4.9.36)

3.2.16.1. GHSTS PDF profile

The GHSTS PDF profile describes some characteristics of the PDF that are recommended for the “Main” file type. It is based on the PDF/A-1 format, as defined in ISO-19005-1:2004. The profile defines additional constraints on top of this base standard and partly relieves constraints from the base standard. The profile is not a complete definition of the file format that allows strict format verification, but a set of recommendations.

The following table defines the GHSTS PDF profile in detail.

	GHSTS PDF profile
Allowed	External references Restriction: Only hyperlinks to computer files that are part of the current submission package in the content folder (see chapter 6.3) are to be used. The hyperlinks in the PDF files should adhere to the relative path definition in chapter 5.2.7. Hyperlinks that refer to individual chapters or pages of a target PDF document are not recommended, as this functionality is not stable.
Recommended	Embedding of all fonts Colour spaces specified in a device-independent manner Inclusion of document structure Language specification Use of metadata
Not recommended	Transparency / overlay Audio / video content Encryption and password protection Digital signatures Embedding of arbitrary file formats Usage of dynamic content (JavaScript) and executable file launches

Table 6 – GHSTS PDF profile

3.2.17. Type of content changes on file level

During the series of submissions of a regulatory action any computer file attached to an instance of FILE, and indirectly DOCUMENT, can change. GHSTS distinguishes two types of content changes:

- Technical changes: Changes on the technical level that have no influence on the content of a computer file, as evaluated by a human being. Examples are:

- Regeneration of a PDF with different settings (e.g. bookmarks).
- Correction of typos on the letter level.
- Change of format for a supplemental file (e.g. using a later Office version).
- Semantic changes: Changes in the content of a computer file as perceived by a human being that may lead to a different evaluation. (e.g. change from a “<” to a “>” sign in a statement about concentrations)

The two types of content changes - “technical change” and “semantic change”- are defined identically for both the “Main” FILE as well as all other optional attachments of a DOCUMENT.

For each FILE instance – used as “Main” or “Attachment” - a “semantic change” triggers the creation of a new PID for the file instance; a “technical change” does not.

Both types of changes will affect the MD5 checksum of the FILE, that is used to verify the correct transmission of a computer file from sender to recipient (see chapter 4.9.36.1).

Technical and semantic changes cannot be distinguished by software. It is the responsibility of the registrant to clearly and carefully distinguish these two types of content changes when setting the values for the content status change of documents that depend on this distinction (see the chapter below).

3.2.18. Content status changes of documents

The DOCUMENT possesses a DOCUMENT_CONTENT_STATUS indicator that is used to track its content changes and subsumes changes on the associated FILE instances.

Let D[x] and D[y] be two DOCUMENT instances from the same document family, with x and y being distinct PID values. The following table summarizes the usage of the content status indicators for DOCUMENT. Please note that the value depends on the distinction of FILE_TYPE values in the categories “Main” and “Attachment” (see chapter 3.2.12)

Value	Description
New	<p>The DOCUMENT content lifecycle status “New” for D[x] is set, if D[x] is new in the dossier, when a “Main” FILE instance is newly present in a submission.</p> <p>When the “Main” file from an existing DOCUMENT instance D[x] needs to be replaced due to a semantic content change, a new DOCUMENT instance D[y] is created as part of the same document family.</p> <p>Notes:</p> <ul style="list-style-type: none"> • • When there is a technical change of the “Main” representation no new DOCUMENT instance with new PID is created – the DOCUMENT instance is not considered to have changed content-wise. • When there is a change to a “attachment” - adding / replacing / content changing (both technical and semantic) / removing - a new DOCUMENT instance is not created. • The PID of the FILE instance does not change in a new DOCUMENT instance if they themselves have not changed due to a semantic change. • For exceptional cases it is permitted to re-introduce a DOCUMENT instance D[x] to a dossier that has been retired before. In this case the content lifecycle status is also “New”.
No change	<p>The composition of the “Main” file and “attachment” files in the DOCUMENT instance remains unchanged since the previous submission.</p> <p>Note: Technical changes may have occurred for “Main” and/or “Attachments”</p>
Modified	<p>There has been a modification in the FILES of FILE_TYPE “Attachments” of the DOCUMENT D[x]: A semantic content change has been applied on one or many attachments, or one or many attachments have been added, removed or replaced. The PID of D[x] and its “Main” File stay the same; the PIDs of the “Attachments” are different compared to the previous submission.</p>
Retired	<p>The DOCUMENT content lifecycle “Retired” for D[x] is set, if the assignment lifecycle status is set to “Retired” for all ToC nodes where D[x] is assigned and is hence retired from the dossier. For assignment lifecycle status see chapter 3.2.20.</p>

Table 7 – document content status indicator

Please note:

- If a document is assigned to ToC node(s) only and is being replaced in the ToC node(s) in the subsequent submission, it will no longer part of the submission package of this submission as each document needs to be assigned to at least one ToC node, and consequently, will neither carry an assignment lifecycle nor a document content lifecycle status.
- Technical changes of the FILE instances of a DOCUMENT do not propagate on the DOCUMENT level – neither the DOCUMENT PID nor the document content status indicator is updated. In case the DOCUMENT has been submitted before, the receiving authority can

compare the MD5 checksum of the FILE instances with those of the previous submission and can decide whether to update their in-house document repository or not.

Each DOCUMENT instance keeps a complete content status history throughout its existence: The element DOCUMENT_CONTENT_STATUS_HISTORY keeps track of the content changes since a DOCUMENT instance with a unique identifier (PID) has been added to the dossier. Each entry in DOCUMENT_CONTENT_STATUS_HISTORY consists of a set of associated values for DOCUMENT_CONTENT_STATUS and SUBMISSION_NUMBER. The values of DOCUMENT_CONTENT_STATUS are defined as a core picklist (see chapter 3.2.24.1).

For the initial addition of the document instance in submission N the element DOCUMENT_CONTENT_STATUS_HISTORY contains the set {New, N}; for each subsequent submission another set of values is added to the element.

3.2.19. Dependencies between content status changes and metadata changes for DOCUMENT

The DOCUMENT concept is the only GHSTS content that possesses elements for metadata changes (see chapter 3.2.3.3) and content changes. For a DOCUMENT added to the dossier both the content status indicator and the metadata status value are set to “New”.

3.2.20. Node assignment status

The node assignment status provides a mechanism to track the assignment of DOCUMENT instances to ToC nodes, i.e. whether a document has been assigned or removed from a particular ToC node. DOCUMENT instances have to be assigned to at least one ToC node and may be assigned to multiple ToC nodes. As ToC nodes represent specific authority requirements, the assignment to more than one ToC node represents the information that a specific document fulfils multiple authority requirements simultaneously.

Once a DOCUMENT is present in a dossier, the node assignment status is independent of the document content status. However, the document content status “Retired” is triggered by the “Retired” assignment status (see chapter 3.2.18).

The following table contains the values for an assignment lifecycle (as represented by the NODE_ASSIGNMENT_STATUS element) regarding a fixed DOCUMENT instance D[x].

Value	Description
New	This DOCUMENT instance D[x] is new in this ToC node
No change	This DOCUMENT instance D[x] is still assigned to this ToC node, compared to the previous submission. In the rare scenario that D[x] is already assigned to this ToC node in the last submission and in parallel replaces another document D[z] in the same ToC node from the same document family in the current submission, then the assignment lifecycle is set to “Replace”.
Replace	This DOCUMENT instance D[x] is set to “Replace”, as D[x] replaces a Document D[z] in the ToC node from the same document family. D[z] is no longer assigned to the ToC node. (Note: The replacing document must be part of the same document family as the replaced document; otherwise, the change has to be modelled as scenario “D[z] = Retired” and “D[x] = New”).

Value	Description
Retired	This DOCUMENT instance D[x] is retired in this ToC node. The document is still considered to be assigned and the status “Retired” is kept for the purpose of lifecycle management for all subsequent submissions.

Table 8 – Node assignment status values

Please note:

- There is a difference between the scenario of retiring a document D[x] and adding D[y] in a ToC node of the same submission versus replacing a document D[x] with a document D[y] in the ToC node: In the first scenario the two DOCUMENT instances have no semantic relations and are from different document families; in the second scenario the two instances are semantically related and must be from the same document family.
- If a DOCUMENT is retired in a ToC node (there is no replacement, meaning the complete document family is retired), then the status assignment status “Retired” is set and the document is kept assigned to the node in the XML backbone (and appears with assignment status “Retired” in the ToC / document list) in all subsequent submissions of the dossier.
- If a DOCUMENT is replaced (with another member of the same document family), then it is removed from the node and it is no longer assigned to the node. This means that there is no assignment lifecycle for the replaced document, but only for the replacing one. One could state that the lifecycle chain continues with the replacing document (with its status “Replace” and the optional reference to the replaced document.).
- A DOCUMENT instance D[x] can only be assigned once to each ToC node.
- A DOCUMENT content lifecycle “New” for a DOCUMENT D[x] can be the result of a node assignment status “New” and/or “Replace”: The content status indicator “New” is also used for a DOCUMENT instance that has replaced a previous DOCUMENT instance of the same document family in a specific ToC node. The DOCUMENT content status would still be “New”, but the node assignment status will be “Replace”.

3.2.21. Differences between content status change and node assignment status

The node assignment lifecycle coexists with the document content lifecycle on the DOCUMENT level. They do not overlap but have a different semantics:

- The node assignment lifecycle is information per ToC node regarding (e.g. assigning, replacing, retiring a DOCUMENT instance to a ToC node) indicating that a specific regulatory information request is satisfied with the assigned document. A document may be assigned to multiple ToC nodes simultaneously and assignments may change over time.
- The document content lifecycle on the DOCUMENT level is information about the composition of the DOCUMENT instance from individual FILE instances, independent of existing assignments to nodes.

3.2.22. Full and incremental submissions

GHSTS supports full and incremental submissions. The following table defines those two terms. They relate to the set of computer files that are referenced by the set of FILE instances.

Submission	Description
Full	The submission package contains all computer files of FILE instances that are part of the GHSTS XML backbone independent of the DOCUMENT_CONTENT_STATUS. The submission package also contains the computer files that are set to “Retired”.
Incremental	The submission package contains only all the computer files of all DOCUMENT instances that have the DOCUMENT_CONTENT_STATUS set to either “New” or “Modified”. "References to FILE instances for DOCUMENT instances where DOCUMENT_CONTENT_STATUS is set to "No change" or "Retired" point either to the submission where the DOCUMENT instance has the latest content change or the last full submission, depending on which is the latest submission.

Table 9 – full and incremental submission

The initial submission of a regulatory action is always full. For each subsequent submission it is up to the communicating parties to decide whether a full or incremental submission should be created.

Please note that the term “incremental submission” only refers to the referenced file content, not to the GHSTS XML backbone. The XML backbone always contains the complete information of the dossier in its current state. The set of FILE and DOCUMENT instances are identical in both cases, only the file system references would differ: In case of a full submission they point to file within the current submission package, for incremental submission they point to a previous submission package.

The only instances that may technically vanish in the GHSTS XML backbone in the course of a regulatory action are FILE or DOCUMENT instances that are replaced by successor instances with new identifiers. However, as usually the replacement is done with similar information (e.g. document from the same document family) the new information is not completely unrelated to the old one.

Computer files already submitted under a different regulatory action should be resubmitted in a full submission; however it is incumbent upon the regulatory authority to decide on the basis of the supplied unique identifiers (see chapter 3.2.26) how to consume it. Therefore the careful assignment of unique identifiers is crucial.

3.2.23. Impact of document content status on evaluation price

The amount of “New” and “Replaced” documents should be a parameter for the effort (costs) of an evaluation of a submission. Documents with status “No Change”, “Modified” or an RA_DOCUMENT_NUMBER should have no impact on the costs.

Therefore industry should use the content status indicators very carefully, for example not declare every document as “New” when it is not, in order to decrease the work of the Regulatory authorities. Each document that is assigned as “New” has to be registered and research has to be done to confirm the document really is new and it has never been reviewed under a different regulatory action with a RA-specific document number.

3.2.24. GHSTS Picklists

GHSTS uses a mix of controlled terminologies and free text values. Where the range of values accepted is controlled a picklist is defined.

Picklists are classified into multiple governance categories:

- CORE: these terms are essential to the specification and are required for the information management functions for example DOCUMENT_CONTENT_STATUS. All GHSTS submissions must respect the CORE picklists.
- Non-CORE:
- DOMAIN: Although the XML backbone is universal, domain picklists allows values to be used to convey information specific to a regulatory domain e.g. PRODUCT_USE.
- RA: picklists which are specific for submissions to a regulatory authority e.g. APPLICATION_TYPE

Where the same terms are used by multiple regulators it is recommended to promote these for DOMAIN picklist and where terms are defined can be used universally it is recommended to promote these to CORE.

The picklists are used by a GHSTS builder to guide the user in the selection of only applicable values. A GHSTS builder is not part of the core components and hence details of how picklists are used is out of scope here. How the resulting GHSTS XML backbone file is filled with values cannot be deduced; it could also be done manually without the aid of a builder and provided picklists provided that the submission passes the technical validation.

Picklists can also be categorized into the following categories:

- When a fixed picklist is used to fill an element, the user has to select one of the predefined picklist entries, with the complete information.
- An open picklist allows the user to supply a decode picklist entry that is not part of the predefined picklist values. In addition to supplying the new decode picklist value, the CODE must be set to the fixed string “other” by the user; this reserved keyword, “other”, must not be pre-allocated in any picklist as CODE.

All picklist values in a GHSTS XML backbone file are set using the following XML type definition:

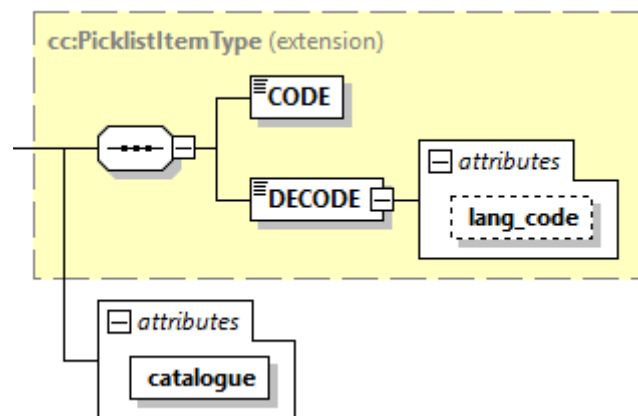


Figure 12 – PicklistItem Type usage

The attribute “catalogue” is an attribute with a specified and fixed value for each picklist-controlled GHSTS element. It serves to indicate which terms are applicable for use with a GHSTS element.

Please note that picklist values in GHSTS have a “flat” and not a hierarchical organisation.

For more details please see the picklist model specification document (see chapter 2.4.1.5).

3.2.24.1. Core GHSTS picklists

The following table provides an overview of the elements in the GHSTS XSD that are controlled by CORE picklists. The values for the core picklists contain the English decode values for explanation. The complete core picklist definition including CODE values is part of the picklist values (see chapter 2.4.1.7)

Concept	ELEMENT in the GHSTS submission XSD	catalogue name (fixed)	English decode Values with explanation
RECEIVER PRODUCT DOCUMENT FILE TOC LEGAL_ENTITY COMPONENT	METADATA_STATUS	metadata_status	Information on whether the metadata for this concept has changed. Values are: <ul style="list-style-type: none"> • New • Modified • No Change Please see chapter 3.2.3.3 for definitions about the respective scope of metadata per concept.
GHSTS	REGULATED_DOMAIN	regulated_domain	Pesticides, GMO, Feed additives, Food contact materials, Food ingredients, Nutrition, Biological hazard
DOCUMENT	DATA_PROTECTION_CLAIM	data_protection_claim	Yes, No, Open
DOCUMENT	DATA_REQUIREMENT	data_requirement	Yes, No, Open
DOCUMENT	DOCUMENT_CONTENT_STATUS	document_content_status	New, Retired, Modified, No Change (see chapter 3.2.18)
LEGAL_ENTITY	COUNTRY	country	List of English country names, with the code being the list of alpha-2 country codes according to ISO 3166-1
COMPONENT	COMPONENT_CLASS	component_class	Organism, Substance (see chapter 3.2.10)
NODE	NODE_ASSIGNMENT_STATUS	node_assignment_status	New, No change, Replace, Retired (see chapter 3.2.20)

Table 10 – Overview of GHSTS elements controlled by core picklists

3.2.24.2. Non-Core GHSTS picklists

The following table provides an overview of the elements in the GHSTS XSD that are controlled by non-core picklists.

Concept	Element	catalogue name (fixed)	Examples of decode values (for explanation only)
PRODUCT	ADMIN_NUMBER_TYPE	admin_number_type	Application Number, BVL Kenn-Nr., EPA Decision No, Experimental Use Permit, Inert Ingredient clearance request
PRODUCT	FORMULATION_TYPE	formulation_type	AE, AL, BB
PRODUCT	PRODUCT_USE	product_use	FEED, FOOD, GREENHOUSE, INDOOR, AGRICULTURE, AMENITY
PRODUCT	PRODUCT_FUNCTION	product_function	Bactericide, Fungicide, Repellent, Novel Food
COMPONENTS	UNIT	unit	units for the components of the product
DOSSIER	REGULATORY_TYPE	regulatory_type	21 CFR Part 11, 91/414/EEC, 1107/2009/EG
DOSSIER	APPLICATION_TYPE	application_type	EU OECD dRR, EPA OPPTS, OECD Annex II Joint Review
DOCUMENT	DOCUMENT_TYPE	document_type	Study report, Study summary
DOCUMENT	RA_DOCUMENT_NUMBER_TYPE	ra_document_number_type	MRID, PMRA document number
DOCUMENT	REFERENCE_TYPE	reference_type	Semantics of reference between documents, E.g. document to submitted document, translation to document

Concept	Element	catalogue name (fixed)	Examples of decode values (for explanation only)
DOCUMENT	DOCUMENT_NUMBER_TYPE	document_number_type	Study Number, Report Number, Project Number, Plot Number
FILE	FILE_TYPE	file_type	Main, Source, Supplemental, (see chapter 3.2.12)
FILE	CBI_DESIGNATION_CLAIM	cbi_designation_claim	yes, no
LEGAL_ENTITY	LEGALENTITY_TYPE	legality_type	Company, Regulatory authority, Consultant, Test House, Third Party
LEGAL_ENTITY	LEGALENTITY_IDENTIFIER_TYPE	legality_identifier_type	DUNS-number, SAP, REACH, VAT-number
COMPONENT	COMPONENT_IDENTIFIER_TYPE	component_identifier_type	CASNO, ECNO, IUBMB
COMPONENT	COMPONENT_DESCRIPTION_TYPE	component_description_type	Event, Species, Strain

Table 11 – Overview of elements controlled by non-core picklists

3.2.25. Regulated Domain

The concept of a “regulated domain” was already introduced in chapter 3.1.1 from a business perspective. From the viewpoint of GHSTS the regulated domain is defined by a specific set of picklist terms that are used to build GHSTS submissions.

In this respect, the selection of a regulated domain (which itself is a core picklist) is a meta-selection and is the first step when creating a new dossier in a GHSTS builder when a new dossier is created as this selection determines the picklist values for all domain-controlled picklists.

Please note that a regulated domain can apply multiple ToC templates. The selection of a regulated domain is not equivalent with the selection of a specific ToC template. The relation between regulated domain and appropriate ToC templates is not represented in GHSTS.

3.2.26. Persistent identifiers

GHSTS enforces the usage of Persistent Identifiers for a number of concepts in GHSTS. The main driver for PID is to avoid duplicate information and to be able to uniquely identify an instance of an information concept throughout its lifetime independent of, for example, variations in spelling.

Most PIDs are created and managed by the registrant. The registrant is responsible for deciding when a new instance of a concept requires a new PID (see chapter 3.2.17 about what is only a modification and what is a new instance). This decision must be done in the scope of the registrant’s internal dossier management system, not in the scope of one individual regulatory action. PIDs identify concepts across different regulatory actions and different recipients, and the same PID has to be used for the same concept across all submissions. In these cases the registrant is required to maintain these PIDs especially in relation to METADATA_STATUS and CONTENT_STATUS updates. The registrant’s inhouse dossier management system must support this lifetime PID management.

Changes in metadata of concepts usually do not change the identity of a concept, and hence the PID can be kept. For example, the change of the common name of a product keeping its formulation or the address details of a legal entity must not trigger a new PID assignment of the respective concept.

The following table lists all elements that have PID identifiers that are managed by the registrant. It also supplies the element names for the concept that serve as user-readable identifiers.

Concept	Element name / description for machine-readable identifier	Element name(s) serving as user-readable identifier
File	FILE_PID (mandatory) Create new value when a new computer file with changed content is stored in the dossier management system of the registrant. No PID must be created for “modifications” (see chapter 3.2.15).	FILENAME (mandatory)
Document	DOCUMENT_PID (mandatory) Create new value when a new document is created in the dossier management system of the registrant (see chapter 3.2.13). A new PID is required exactly when the PID of the “Main” content file is new or has changed.	DOCUMENT_NUMBER (0..n)
Document family	DOCUMENT_FAMILY_PID (mandatory) Create new value when the initial document for a specific information (e.g. a specific study report) is created in the dossier management system of the registrant. Subsequent variations or versions of the same information will carry the same DOCUMENT_FAMILY_PID.	DOCUMENT_FAMILY (mandatory)
Dossier	DOSSIER_PID (mandatory) Create new value when new regulatory action is created in the dossier management system of the registrant.	DOSSIER_DESCRIPTION_TITLE (mandatory), PROJECT_ID_NUMBER (0..n)
Product	PRODUCT_PID (mandatory) For the pesticide domain: A “product” is a Product or a Technical Grade Active Ingredient (U.S.)/Active Substance (EU). A new Product PID has to be generated when its composition of components is changed.	GENERIC_PRODUCT_NAME (optional) PRODUCT_NAME (optional), ADMIN_NUMBER (optional)
Legal entity	LEGALENTITY_PID (mandatory) Unique identifiers for registrant and regulatory authority. If no PID value is available in a central repository, then the usage of the “nil PID” is permitted (see chapter 5.2.9)..	LEGALENTITY_NAME (mandatory), OTHER_NAME (0..n)
Component	COMPONENT_PID (optional) The same PID should be used if different formulations or products from the registrant contain the same component. Unique identifier for a component. Currently a similar PID is created by IUCLID using the UUID format (see chapter 5.2.9.4).	COMPONENT_NAME (mandatory), COMPONENT_IDENTIFIER (0..n) COMPONENT_DESCRIPTION (0..n)

Table 12 – Overview of concepts identified by PID generated by the registrant

For two concepts in the GHSTS submission that need to be uniquely identified across registrants and authorities, the PIDs and their associated information concepts are generated and supplied centrally; every registrant needs to use those PIDs and their associated information in the submission: PIDs in standard ToC templates (TOC_PID and TOC_NODE_PID) – see chapter 2.4.2.5.

Concept	Element name / description for machine-readable identifier	Element name(s) serving as user-readable identifier
ToC	TOC_PID (mandatory) Identification of a standard ToC template. If values are provided the registrant should use them throughout the regulatory action.	TOC_SHORT_NAME (mandatory), TOC_FULL_NAME (mandatory)
ToC node	TOC_NODE_PID (mandatory) Identification of a node within a standard ToC. If values are provided the registrant should use them throughout the regulatory action.	TOC_NODE_NAME (mandatory)

Table 13 – Overview of concepts identified centrally managed PID

The registrant generates and maintains all other PID values which are to be immutable throughout their lifecycle, also across divestments (e.g. the same PID for the same product, the same PID for the legal entity for all submissions it receives).

For the description how a PID is technically represented please see chapter 5.2.9.

3.2.27. Further identifiers on submitter and regulatory authority level

Elements of GHSTS can be used to transfer other identifiers such as those issued by the regulatory authority or the testing laboratory. The following table provides an overview.

Concept	Stakeholder	Element name / description
Regulatory Action	Regulatory authority	PROJECT_ID_NUMBER Each participating regulatory authority can assign a Project ID number to a dossier. Recommendation: <ul style="list-style-type: none"> Use this element to allow authorities to group multiple dossiers, e.g. AI plus formulation dossiers It is up to authority to decide upon the semantics and grouping scope of the PROJECT_ID_NUMBER and communicate this to the submitting party
Product	Regulatory authority	ADMIN_NUMBER Each participating regulatory authority can assign a regulatory action specific admin number to a product. The type can be specified (BVL Kenn-Nr., EPA Decision No, ...).
Document	Regulatory authority	RA_DOCUMENT_NUMBER Each participating regulatory authority can assign a regulatory authority specific Document number to the Document. The type can be specified (MRID, PRMA document number, ...).
Document	Registrant	DOCUMENT_NUMBER The registrant can assign one or many different document numbers and specify their individual type (Report number, Study number, Project number, Plot number, Old report number, other).

Table 14 – Legal entity specific identifier

3.2.28. Information for and targeting at individual receivers

GHSTS allows the addition of metadata for individual regulatory authorities (RA) acting as receivers. The following table summarizes the RA-specific elements and the information that can be delivered with the element and its sub-elements. The column “Man” lists, whether the information is mandatory to be provided for *all* receivers of the submission.

Concept	Element	Man for all receivers	Description of RA-specific information
PRODUCT	PRODUCT_RA	N	Allows to specify an RA-specific product name and an ADMIN_NUMBER (see table “Table 14 – Legal entity specific identifier”)
DOSSIER	DOSSIER_RA	Y	Specification of the regulatory type, the application type (for both see “Table 11 – Overview of elements controlled by non-core picklists”) and a RA-specific Project ID number (see Table 14 – Legal entity specific identifier).
DOCUMENT	DOCUMENT_RA	N	If element is present, then this document is targeted at this RA. A document must be targeted at a minimum to one receiving RA. Specification of additional metadata: data protection, the data requirement (for both see “Table 10 – Overview of GHSTS elements controlled by core picklists”), a document comment, other national guideline number and guideline systems and an RA-specific document number (see table “Table 14 – Legal entity specific identifier”)
FILE	FILE_RA	N	If element is present, then this file is targeted at this RA. A file must be targeted at a minimum to one receiving RA. Specification of additional metadata: CBI designation (CBI = Confidential Business Information) and a file comment

Table 15 – Overview of RA-specific information

The elements DOCUMENT_RA and FILE_RA are used both for targeting information at RA as well as providing additional RA-specific information. If there is no element DOCUMENT_RA for a specific RA, then this document and all referenced FILE instances are not relevant and can be ignored by this RA. If there is an element DOCUMENT_RA for a specific RA, but for parts of the FILE instances of this document there is no element FILE_RA for the same RA, then those referenced files can be ignored by the RA.

Each RA-specific element has an XML attribute @To_Specific_for_RA_Id to make references to the corresponding attribute “Id” of the receiver element (RECEIVER@Id).

The following figure depicts the references in a graphical manner.

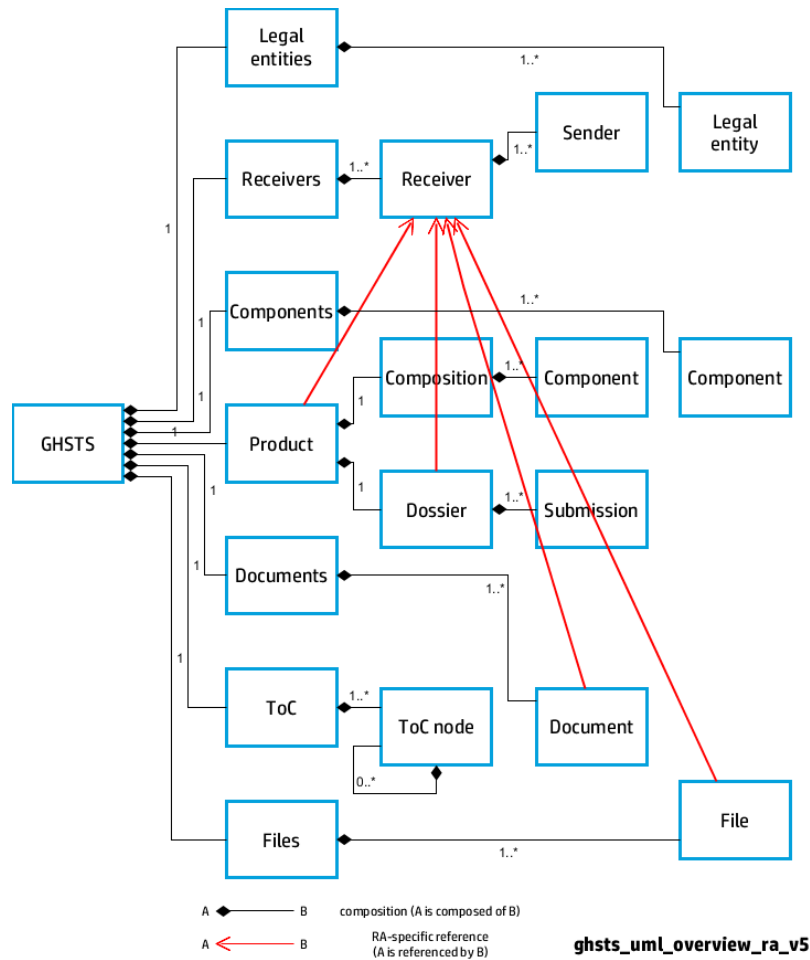


Figure 13 – overview of receiver-specific references

Please note that the information in the RA-elements serves as *additional* RA-specific information for the respective instances of the concept. Though GHSTS supports assigning instances of concepts (e.g. products, files, documents) for the use of specific receivers (RA), all receivers receive the same complete submission package and any receiver can access also the information that is targeted for other RA.

3.2.29. Internal and external references

The following table shows all references that can explicitly be set in the GHSTS backbone. Internal references have a target concept within the same submission package; external references have the target concept outside of the same submission package (in a previous submission of the same dossier or outside of the dossier).

Element to represent reference (incl. cardinality)	From concept	To concept	int ext	Description
REFERENCED_DOSSIER (0..n)	Dossier	Dossier	ext	Reference from a current dossier to another referenced dossier. The referenced dossier is identified with its PID or a previous dossier numbering system for example the Caddy Dossier ID (e.g. LANDE005). A reference reason can be provided. The referenced dossier can be outside of the current regulatory action.

Element to represent reference (incl. cardinality)	From concept	To concept	int ext	Description
DOSSIER_RA (1..n)	Dossier	Receiver	int	Reference from Dossier to Receiver to specify for which RA the additional metadata below DOSSIER_RA is assigned. The target Receiver is identified with unique key in attribute To_Specific_for_RA_Id.
REFERENCED_DOCUMENT (0..n)	Document	Document	int ext	Reference between documents, both to model relations between changed documents during document evolution in a document family as well as other generic types of references. The referenced document can be internal (within the current dossier and its submissions) or external (outside of the current dossier). The referenced document is identified by its PID, in the case of external references also existing document numbers can be used, if no document PID has been assigned and is known to the receivers. External references can be inside or outside of the current regulatory action.
RELATED_TO_COMPONENT (0..n)	Document	Component	int	Reference from Document to component to specify relationship of document with component. The target component is identified with unique key in attribute To_Component_Id.
DOCUMENT_RA (1..n)	Document	Receiver	int	Reference from Document to Receiver to specify for which RA the document is targeted at and to provide additional metadata. The target Receiver is identified with unique key in attribute To_Specific_for_RA_Id.
REFERENCED_TO_FILE (1..n)	Document	File	int	Reference from document to all file instances of which the document is composed of. File is identified with unique key stored in attribute To_File_Id.
FILENAME (1..1)	File	File system	int	Reference from file to file system. The element stores a relative file path (for details see chapter 5.2.7)
FILE_RA (1..n)	File	Receiver	int	Reference from File to Receiver to specify for which RA the file is targeted and to provide additional metadata. The target Receiver is identified with unique key in attribute To_Specific_for_RA_Id.
TOC2DOC (0..n)	ToC Node	Document	int	Reference from ToC node to document. The document is identified with unique key stored in attribute To_Document_Id.
STANDARD_TOC_REFERENCE (0..1)	ToC	Standard ToC	int	Reference to standard ToC XML file with PID and filename (the standard ToC is an optional separate XML file in the submission package)

Element to represent reference (incl. cardinality)	From concept	To concept	int ext	Description
RECEIVER (1..n)	Receiver	Legal Entity	int	Reference from Receiver to Legal Entity (Legal Entity has standard contact information, Receiver adds specific information). Legal Entity is identified via unique key in attribute To_Legal_Entity_Id.
SENDER (1..n)	Sender	Legal Entity	int	Reference from Sender to Legal Entity (Legal Entity has standard contact information, Receiver adds specific information). Legal Entity is identified via unique key in attribute To_Legal_Entity_Id.
PRODUCT_RA (0..n)	Product	Receiver	int	Reference from Product to Receiver to specify for which RA the additional metadata below PRODUCT_RA is assigned. The target Receiver is identified with unique key in attribute To_Specific_for_RA_Id.
COMPOSITION (1..n)	Component	Component	int	Reference from Component (below PRODUCT) to components to specify for which component the quantity in the product composition is provided. The target component is identified with unique key in attribute To_Component_Id.

Table 16 – Overview of references

All references that are exclusively internal within the same GHSTS XML backbone file use a unique key to identify the target object using XML methods: The identifier is of type XS:ID and the reference to the identifier of type XS:IDREF. This unique key is deliberately different from the PID concept.

The internal references are depicted in “**Figure 8 –Main GHSTS concepts and their relations**”, the external references are depicted in the following figure.

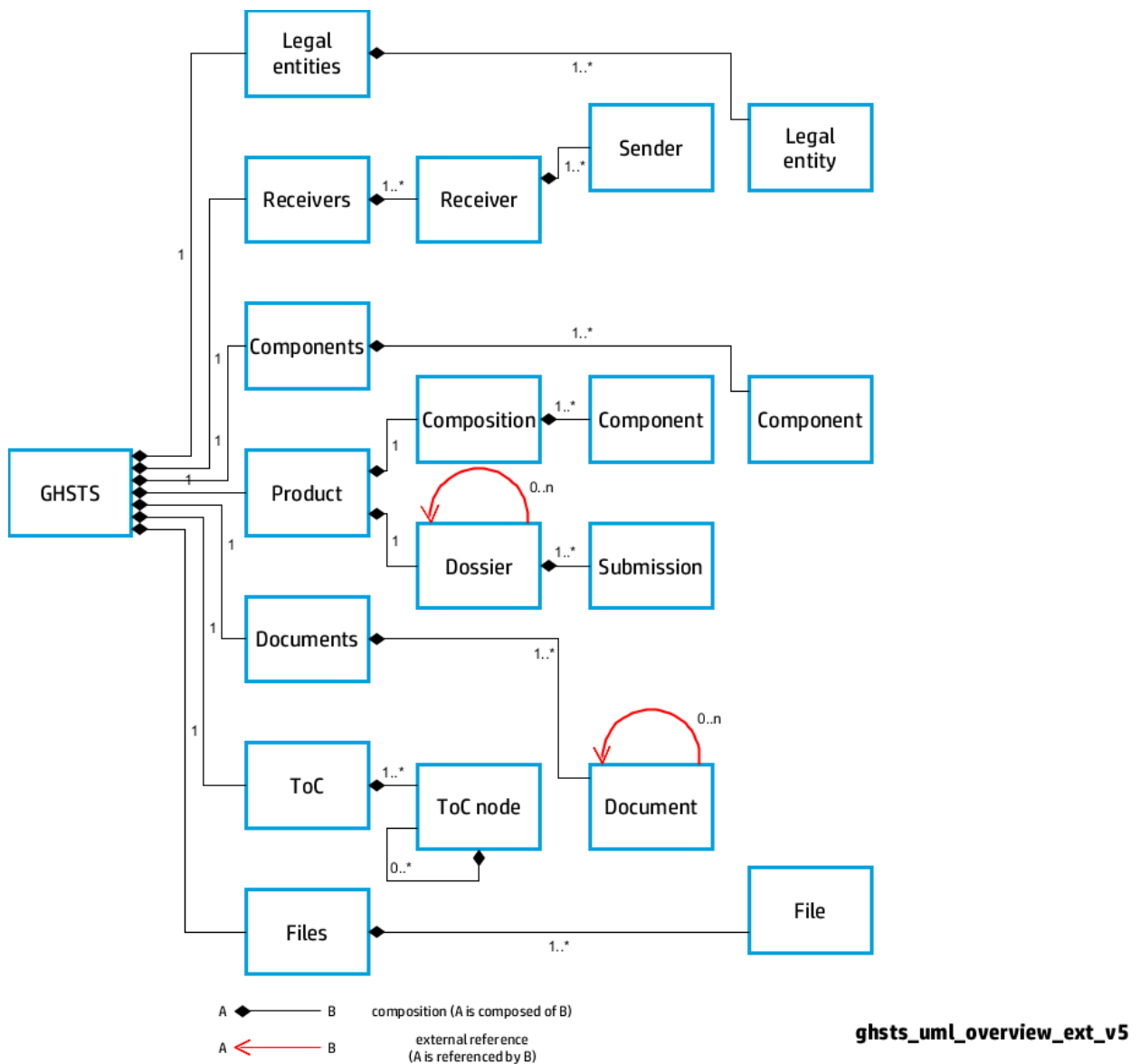


Figure 14 – overview of external references

The figure shows with red arrows all references that can point from the element of one submission to the corresponding element of a predecessor GHSTS submission (for Document and File) or to an arbitrary GHSTS or non-GHSTS submission (for Dossier and Document) to place the information in a larger context (i.e. the referenced elements are not considered part of the submission from which the elements are referenced). For example, a product dossier registration as regulatory action may reference the preceding substance registration.

Document identifiers (e.g. DOI) could also be considered as external references, they are not part of the above figure.

Chapter 4. GHSTS schema definition

4.1. Overview

This chapter provides a technical reference for the GHSTS XML schema definition, as defined in the GHSTS XSD schema file. This chapter contains the most relevant aspects of the XSD, especially an explanation on how elements should be used in the business context. The complete definition is contained in the XSD itself.

This chapter is intended for

- IT project leads and IT consultants that need to understand how to generate a valid GHSTS XML backbone file. For this audience knowledge of XML, XSD and related technologies (e.g. XPath) is recommended.
- Business domain experts that need to know the semantics of individual elements of the GHSTS metadata. This audience can skip the technical aspects of this chapter and directly address the description of the individual elements.

4.2. Filename and version

The GHSTS XSD is versioned using a version numbering scheme of type XX.XX.XX. It is composed of a two digit major, a two digit minor and a two digit patch level number separated by single dots. Major versions are used for global functional changes, minor versions for local changes and patch level versions for error corrections. Patch level versions do not require new versions of other GHSTS components (see chapter 2.4).

This version is contained in the version attribute of the root schema element in the XSD file.

The filename uses the same versioning scheme with the hyphen instead of the dot: The file naming convention of the GHSTS XSD is therefore

ghsts_XX-XX-XX.xsd

for example

ghsts_01-00-00.xsd

This chapter describes the version 02.00.00 of the GHSTS XML schema, published on <http://www.oecd.org/chemicalsafety/submission-transport-standard/>.

4.3. Scope

The GHSTS XML schema definition describes the structure of the XML backbone file of *one* single submission that is part of a regulatory action. The XML backbone file has the following references to the outside world:

- For internal and external references from the XML backbone see chapter 3.2.28.

- For the folder structure of the submission package of which the XML backbone is part of, and the folder structure of the parent regulatory action see chapter Chapter 6.

4.4. Namespace

The GHSTS schema uses the following namespace definitions:

- xmlns:xs="http://www.w3.org/2001/XMLSchema"
- xmlns="http://www.oecd.org/GHSTS"
- targetNamespace="http://www.oecd.org/GHSTS"

4.5. Unique keys

As described in chapter 3.2.29, the internal references within a submission are made using unique keys to identify the target object. The unique key of the target object is represented as attribute "Id" and is of type XS:ID, the reference to the target object is represented as attribute "To_<xxx>" with xxx being a variable string. The reference is of type XS:IDREF.

Please note that values of the attribute "Id" must be valid XML names, which means that they must start with a character. For details please consult the definition of the XML type XS:ID.

The following table lists the XPath expression of the attributes with Id values as well as the XPath expression of the attributes that reference this unique key.

XPath of element with Id attribute	Xpath of element attribute with reference to element in first column
GHSTS/LEGAL_ENTITIES/LEGAL_ENTITY/@Id	GHSTS/PRODUCT/DOSSIER/DOSSIER_RA/@To_Specific_for_RA_Id GHSTS/DOCUMENTS/DOCUMENT/DOCUMENT_RA/ @To_Specific_for_RA_Id GHSTS/FILES/FILE/FILE_RA/@To_Specific_for_RA_Id GHSTS/PRODUCT/PRODUCT_RA/@To_Specific_for_RA_Id GHSTS/RECEIVERS/RECEIVER/@To_Legal_Entity_Id GHSTS/RECEIVERS/RECEIVER/SENDER/@To_Legal_Entity_Id
GHSTS/COMPONENTS/COMPONENT/@Id	GHSTS/DOCUMENTS/DOCUMENT/DOCUMENT_GENERIC/RELATED_TO_COMPONENT/ @To_Component_Id GHSTS/PRODUCT/COMPOSITION/COMPONENT/@To_Component_Id
GHSTS/FILES/FILE/@Id	GHSTS/DOCUMENTS/DOCUMENT/DOCUMENT_GENERIC/REFERENCED_TO_FILE/@To_File_Id
GHSTS/DOCUMENTS/DOCUMENT/@Id	GHSTS/TOC/STRUCTURE//TOC_NODE/TOC2DOC/@To_Document_Id

Table 17 – internal references with unique keys

The value for the unique key can be freely assigned by the registrant during the publication process. The only constraints are the uniqueness for each concept and for each submission package. The GHSTS XSD contains checks to validate the uniqueness constraint and the referential integrity of internal references by checking the existence of the target element with the unique key.

4.6. Combination of value and value type

There are a number of XML elements of the pattern "value type / value", where the value type is picklist-controlled. (e.g. DOCUMENT_NUMBER_TYPE, COMPONENT_DESCRIPTION_TYPE). Those elements have the following type of structure (with DOCUMENT_NUMBER as example)

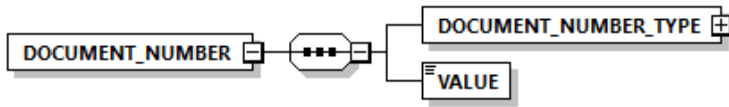


Figure 15 – value and value type

For example, a specific document number (e.g. X120976) will be submitted using the VALUE element in combination with the corresponding DOCUMENT_NUMBER_TYPE, which characterizes the type of the number using a picklist (e.g. “Report Number”). A VALUE can never be supplied without a corresponding type element.

This structure is the same for several other elements (see table below). The name of the element for the type of the value is derived from the root element name:

Root element	Name of type element
ADMIN_NUMBER	ADMIN_NUMBER_TYPE
DOCUMENT_NUMBER	DOCUMENT_NUMBER_TYPE
LEGALENTITY_IDENTIFIER	LEGALENTITY_IDENTIFIER_TYPE
RA_DOCUMENT_NUMBER	RA_DOCUMENT_NUMBER_TYPE
COMPONENT_IDENTIFIER	COMPONENT_IDENTIFIER_TYPE
COMPONENT_DESCRIPTION	COMPONENT_DESCRIPTION_TYPE

Table 18 – GHSTS elements with “value type / value” pattern

4.7. Notation conventions

This chapter provides a brief overview about the notation conventions that are used to explain the schema definition as well as conventions used for the creation of the schema. It serves to support readers from the business audience to understand the technical core characteristics of the schema.

4.7.1. Elements

The element names of the Schema Description File are uppercase strings. If an element name consists of multiple words, they are concatenated by an underscore.

```
<xs:element name="DOSSIER_DESCRIPTION_TITLE">
```

The principle of the GHSTS schema is to transport element values in the GHSTS XML File as text content of the elements.

```
<DOSSIER_DESCRIPTION_TITLE>The text content</DOSSIER_DESCRIPTION_TITLE>
```

4.7.2. Attributes

The Attribute names are mixed uppercase and lower case strings e.g. “To_Legal_Entity_Id“. Most attribute values are used for technical aspects in a GHSTS XML file to characterize nodes with unique identifiers or to make references to these identifiers.

```
...<FILE Id="The attribute value"> ...</FILE>
```

4.7.3. Occurrence indicators

The minOccurs / maxOccurs indicator specifies the minimum / maximum number of times an element can occur. For mandatory values the value for minOccurs must be always one, as here for the definition of the number of files that a document must possess, it is a mandatory multi-value element:

```
<xs:element name="REFERENCED_TO_FILE" minOccurs="1" maxOccurs="unbounded">
```

The following table shows how the occurrence indicators are displayed in the graphical schema representation, using elements from the DOCUMENT concept.

Occurrence	Occurrence indicators	Graphical representation
mandatory / single value (1/1)	minOccurs="1" maxOccurs="1"	
optional / single value (0/1)	minOccurs="0" maxOccurs="1"	
mandatory / multi-value (1/*)	minOccurs="1" maxOccurs="unbounded"	
optional / multi-value (0/*)	minOccurs="0" maxOccurs="unbounded"	

Table 19 – Notation and graphical representation of occurrence indicators

Please note that if not stated otherwise, the order of multiple instances within the XML file is of no business significance. Two apparent exceptions are the order of ToC nodes and the order of document references in a ToC node.

4.8. Boilerplate for element description

This chapter contains the boilerplate for the subsequent documentation of the XSD elements including an explanation of the description template.

Please note that this specification contains only main aspects of the actual XSD definition. The full XSD definition can be found in the XSD file itself

4.8.1. Child elements

This chapter defines the child elements of the element in focus in the form of a table:

Element name				
Xpath				
Diagram				
Child element	T	Occ	Definition	Description

Table 20 – boilerplate sample: child elements definition

- The row Xpath contains the Xpath expression of the element in an instance.
- The cell next to “Diagram” shows a graphical representation of the element and its direct child elements.
- The rows below “child elements” contain the child elements of the element in focus. The order of child elements in the parent element is of no significance.
- The column “T” stands for “Type”. It contains the character “T” if the child element contains text content, and the character “E” if the element itself contains element content.
- The column “Occ” contains the occurrence indicator as explained in chapter 4.7.3, e.g. “(0/1)”.
- The column “Definition” contains the definition of the element including a reference where the definition can be found (except for standard XML schema “xs” types).
- For GHSTS-specific types the column “definition” contains a reference to the chapter with the definition of the element.

4.8.2. Attributes

For elements with attributes, this chapter lists the attributes of the element:

Attribute name	Man	Type	Definition	Description

Table 21 - boilerplate sample: attributes definition

- The column “Man” specifies whether the attribute is mandatory or not (Y = Yes, N = No).

The chapter is omitted when an element does not contain any attributes.

4.8.3. Constraints

The chapter “Constraints” includes additional constraints and guidance on how to use the element and its child elements and attributes.

Constraint Name	Description

Table 22 - boilerplate sample: constraints definition

This chapter may serve as input for the specification of the GHSTS validator. Please note that the GHSTS validator may perform additional checks on the schema not described here.

The chapter is omitted when an element does not contain any additional constraints.

4.8.4. Example

This chapter contains a short XML example of an element instance and its direct child elements, omitting the further nested XML elements.

Note: This chapter does not appear for all element instances and child elements.

4.9. Elements

4.9.1. Overview

An overview diagram of the complete XSD is provided as separate image file in PNG format, with the filename

ghsts_XX-XX-XX.png

with XX-XX-XX being the version string of the current GHSTS specification version (see chapter 1.4).

4.9.2. GHSTS

The GHSTS element is the root element of the complete XML backbone file.

4.9.2.1. Child elements

Element name	GHSTS			
Xpath	GHSTS			
Diagram				
Child element	T	Occ	Definition	Description
REGULATED_DOMAIN	E	(1,1)	chapter 4.9.3	Regulated domain of the submission
SENDER_ENVIRONMENT	E	(0,1)	chapter 4.9.4	Technical build and validation information of the sender
RECEIVERS	E	(1,1)	chapter 4.9.3	List of receivers
PRODUCT	E	(1,1)	chapter 4.9.10	Product
COMPONENTS	E	(1,1)	chapter 4.9.20	List of components
DOCUMENTS	E	(1,1)	chapter 4.9.18	List of documents
FILES	E	(1,1)	chapter 4.9.33	List of files

TOC	E	(1,1)	chapter 4.9.37	ToC
LEGAL_ENTITIES	E	(1,1)	chapter 4.9.40	List of legal entities

Table 23 –child elements definition for element GHSTS

4.9.2.2. Attributes

Attribute name	Man	Type	Definition	Description
specificationversion	Y	versionnumberType	chapter 4.2	Version of the GHSTS specification (this document) to which this submission adheres. Please note that the version of the GHSTS specification is not identical to the version of the GHSTS submission XSD. The GHSTS XSD is referenced in the standard attribute schemaLocation (see example below)
xsi:schemaLocation	Y		chapter 8.2 XML schema	reference to the GHSTS submission XSD file, see example below. The location of the XSD and the path has to be compliant with the folder structure, see chapter 6.1.
xmlns	Y		chapter 8.2 XML schema	definition of the standard GHSTS namespace, see example below. Do not change this value.
xmlns:xsi	Y		chapter 8.2 XML schema	definition of the standard XSI namespace (XML schema instance). Do not change this value.

Table 24 - attributes definition for element GHSTS

4.9.2.3. Example

```
<GHSTS specificationversion="02.00.00" xsi:schemaLocation="http://www.oecd.org/GHSTS
../02/utls/ghsts_02-00-00.xsd" xmlns="http://www.oecd.org/GHSTS"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
```

4.9.3. REGULATED_DOMAIN

The element contains the information about the regulated domain of the regulatory action of the dossier and its submission. The picklist values used must be compliant with CORE and the relevant DOMAIN lists. The value of the regulated domain must be maintained in all submissions of the same dossier.

4.9.4. SENDER_ENVIRONMENT

This element combines information about the optional build and validation information. If this element is present, either build or validation information must not be empty.

4.9.4.1. Child elements

Element name	SENDER_ENVIRONMENT			
Xpath	GHSTS/SENDER_ENVIRONMENT			
Diagram	<p>The diagram illustrates the relationship between the SENDER_ENVIRONMENT element and its child elements. SENDER_ENVIRONMENT is shown as a dashed box containing a solid box with a horizontal line and a right-pointing arrow. This solid box is connected to a solid box with a horizontal line and a right-pointing arrow, which in turn is connected to a dashed box containing two solid boxes, one labeled BUILDER and one labeled VALIDATOR, both with horizontal lines and right-pointing arrows.</p>			
Child element	T	Occ	Definition	Description
BUILDER	E	(* ,1)	Chapter 4.9.5	Build information
VALIDATOR	E	(* ,1)	Chapter 4.9.6	Latest validation information

Table 25 –child elements definition for element SENDER_ENVIRONMENT

4.9.5. BUILDER

This optional freetext XML element allows to capture information about the build environment, build date and release and version date of applied picklist values that were used to create the present submission. The builder tool / submitter can decide in detail what to supply. This element is intended to be used for support information. It must not be used for regulatory processes or automated workflows that influence GHSTS processing.

4.9.5.1. Attributes

Attribute name	Man	Type	Definition	Description
build_date	Y	xs:date		Date of technical building the present submission package, may be different from the official SUBMISSION_VERSION_DATE.

Table 26 - attributes definition for element BUILDER

4.9.6. VALIDATOR

This optional freetext XML elements allows to capture information about the latest applied validation. The information that can be added is up to the provider of the validator but should include brief information about applied GHSTS validation rule set (as defined in future by EGEEPD) and brief validation results. This information does not replace a validation report of the validator itself (partly requested by authorities e.g. for CADDY). but is intended for support use cases. It must not be used for regulatory processes or automated workflows that influence GHSTS processing.

4.9.6.1. Attributes

Attribute name	Man	Type	Definition	Description
validation_date	Y	xs:date		Date of last pre-submission validation of the present submission package, may be different from the official SUBMISSION_VERSION_DATE.

Table 27 - attributes definition for element BUILDER

4.9.7. RECEIVERS

The RECEIVERS element is the definition block for all Regulatory authorities who will be a receiver of the GHSTS Submission Package and the relationship to the corresponding SENDER (Submitting Companies). This multi valued relationship is needed for Joint Review.

4.9.7.1. Child elements

Element name	RECEIVERS			
Xpath	GHSTS/RECEIVERS			
Diagram				
Child element	T	Occ	Definition	Description
RECEIVER	E	(1,*)	chapter 4.9.8	Single receiver.

Table 28 –child elements definition for element RECEIVERS

4.9.8. RECEIVER

The element describes a single receiver of a submission. The main information about the receiver is provided by the element LEGAL_ENTITY which is referenced in an attribute.

For the explanation of the logical concept please see chapter 3.2.8.

4.9.8.1. Child elements

Element name	RECEIVER			
Xpath	GHSTS/RECEIVERS/RECEIVER			
Diagram				
Child element	T	Occ	Definition	Description
METADATA_STATUS	E	(1,1)	picklist, see chapter 3.2.24	see chapter 3.2.3
SHORT_NAME	T	(1,1)	xs:string	Typical abbreviation of the receiver, e.g. EPA, BVL. Used for display in the GHSTS browser.
ROLE	T	(0,1)	xs:string	Each receiver is characterized by his role in the work sharing process. Example: “RMS”

SENDER	E	(1,*)	chapter 4.9.9	For each receiver multiple senders can be supplied.
--------	---	-------	---------------	---

Table 29 –child elements definition for element REICEIVER

4.9.8.2.Attributes

Attribute name	Man	Type	Definition	Description
Id	Y	XS:ID	chapter 4.5	unique ID, which is used for reference in other elements that are specific for a RA (attribute To_Specific_for_RA_Id)
To_Legal_Entity_Id	Y	XS:IDREF	chapter 4.5	ID reference to legal entity instance supplying the details for the receiver. Reference validated in schema using keyref RefIntegrityReceiverToLegalentity

Table 30 - attributes definition for element RECEIVER

4.9.9. SENDER

The element SENDER describes a sender of a submission. Please note that this information is RA-specific: For each RECEIVER different and eventually multiple senders can be submitted.

The main information about the sender is provided by the element LEGAL_ENTITY which is referenced in an attribute. For the explanation of the logical concept please see chapter 3.2.8.

4.9.9.1.Child elements

Element name	SENDER			
Xpath	GHSTS/RECEIVERS/RECEIVER/SENDER			
Diagram				
Child element	T	Occ	Definition	Description
COMPANY_CONTACT_REGULATORY_ROLE	T	(0,1)	xs:string	Each sender is characterized by his role in the work sharing process. In case of “joint submissions” there should be defined one company with the role “Lead submitting company”. That means that all additional submissions were built by this company.
REMARK	T	(0,1)	xs:string	Area to communicate additional company contact details.

Table 31 –child elements definition for element SENDER

4.9.9.2.Attributes

Attribute name	Man	Type	Definition	Description
To_Legal_Entity_Id	Y	XS:IDREF	chapter 4.5	ID reference to legal entity instance supplying the details for the sender. Reference validated in schema using keyref RefIntegritySenderToLegalentity

Table 32 - attributes definition for element SENDER

4.9.10. PRODUCT

The element PRODUCT represents the product that is the subject of the complete regulatory action. It contains the DOSSIER and SUBMISSION as direct and indirect child elements, because both are dependent on the PRODUCT.

There is exactly one product per regulatory action.

For the explanation of the logical concept please see chapter 3.2.9.

4.9.10.1.Child elements

Element name	PRODUCT			
Xpath	GHSTS/PRODUCT			
Diagram				
Child element	T	Occ	Definition	Description
METADATA_STATUS	E	(1,1)	picklist, see chapter 3.2.24	picklist, see chapter 3.2.3
PRODUCT_PID	T	(1,1)	PIDType, chapter 5.2.9	unique PID of the product, see chapter 3.2.26.
GENERIC_PRODUCT_NAME	T	(1,1)	xs:string	generic name (common name); In the pesticide domain also the active ingredient

PRODUCT_RA	E	(0,*)	chapter 4.9.11	optional element to supply additional metadata for each receiver.
FORMULATION_TYPE	E	(0,1)	picklist, see chapter 3.2.24	picklist, see chapter 3.2.3 Two character code describing the formulation of the product following GIFAP rules, according to "Catalogue of Pesticide Formulation types and International Coding Systems " GIFAP Technical Monograph No. 2 4th Edition 1999 Examples: "EC", "WP", "GR"
COMPOSITION	E	(1,1)	Chapter 4.9.12	list of components that constitute the product
PRODUCT_USE	E		picklist, see chapter 3.2.24	The proposed use of the product, e.g. FEED, FOOD, GREENHOUSE, INDOOR, AGRICULTURE, AMENITY. The usage of the product defines the populations which may be exposed to the product and as a consequence the type of evidence (in the form of documents and data) required for the risk assessment.
PRODUCT_FUNCTION	E		picklist, see chapter 3.2.24	function of the product
DOSSIER	E	(1,1)	chapter 4.9.14	dossier element as representation of the metadata for the complete regulatory action = sequence of GHSTS submissions

Table 33 –child elements definition for element PRODUCT

4.9.11. PRODUCT_RA

For the explanation of the logical concept please see chapter 3.2.28.

4.9.11.1. Child elements

Element name	PRODUCT_RA			
Xpath	GHSTS/PRODUCT/PRODUCT_RA			
Diagram				
Child element	T	Occ	Definition	Description
PRODUCT_NAME	T	(0,1)	xs:string	RA-specific product name associated with the dossier

ADMIN_NUMBER	E	(0,*)	chapter 4.6	product references to the numbering world of a Regulatory authority. Examples: “BVL Kenn-Nr.”, “EPA Decision No”, “PMRA Decision No”, “PRIA Pre-application”.
--------------	---	-------	-------------	--

Table 34 –child elements definition for element PRODUCT_RA

4.9.11.2.Attributes

Attribute name	Man	Type	Definition	Description
To_Specific_for_RA_Id	Y	XS:IDREF	chapter 4.5	ID reference to legal entity instance. Reference validated in schema using keyref RefIntegrityProduct_RAToLegalentity

Table 35 - attributes definition for element PRODUCT_RA

4.9.11.3.Constraints

Constraint Name	Description
	Elements PRODUCT_NAME and ADMIN_NUMBER must not both be missing, at least one element has to be present.

Table 36 - constraints definition for element PRODUCT_RA

4.9.12. COMPOSITION

This element contains the list of components in the product.

4.9.12.1.Child elements

Element name	COMPOSITION			
Xpath	GHSTS/PRODUCT/COMPOSITION			
Diagram				
Child element	T	Occ	Definition	Description
COMPONENT	E	(1,*)	Chapter 4.9.13	single component in the product

Table 37 –child elements definition for element COMPOSITION

4.9.13. COMPONENT

The information in the element COMPONENT provides the definition of the quantity of the respective component in the product. The component specific metadata are submitted in the element GHSTS/COMPONENTS/COMPONENT, which is referred by attribute To_Component_Id.

For the explanation of the logical concept please see chapter 3.2.9.

4.9.13.1. Child elements

Element name	COMPONENT			
Xpath	GHSTS/PRODUCT/COMPOSITION/COMPONENT			
Diagram				
Child element	T	Occ	Definition	Description
QUANTITY	T	(0,1)	xs:decimal	Quantity, specified as decimal Example: “42.729”
UNIT	E	(0,1)	picklist, see chapter 3.2.24	Example: “g/kg”

Table 38 –child elements definition for element COMPOSITION

4.9.13.2. Attributes

Attribute name	Man	Type	Definition	Description
To_Component_Id	Y	XS:IDREF	chapter 4.5	ID reference to component instance in GHSTS/COMPONENTS/COMPONENT. Reference validated in schema using keyref RefIntegrityCompositionToComponent

Table 39 - attributes definition for element COMPOSITION

4.9.13.3. Constraints

Constraint Name	Description
	For the pesticide domain: For TGAI / substance dossiers both quantity and unit can be empty. For formulated product dossiers both values have to be supplied.

Table 40 - constraints definition for element COMPONENT below PRODUCT

4.9.14. DOSSIER

Representation of the metadata for the dossier for one complete regulatory action as ordered sequence of GHSTS submission.

For the explanation of the logical concept please see chapter 3.2.6.

4.9.14.1. Child elements

Element name	DOSSIER			
Xpath	GHSTS/PRODUCT/DOSSIER			
Diagram				
Child element	T	Occ	Definition	Description
DOSSIER_PID	T	(1,1)	PIDType, chapter 5.2.9	unique PID of the dossier, see chapter 3.2.26.
DOSSIER_DESCRIPTION_TITLE	T	(1,1)	xs:string	human-readable identifier of the dossier designated by the registrant. Examples: “OECD Annex II Dossier Spiroxamine”, “ACME Natural Insecticide”
DOSSIER_COMP_ID	T	(1,1)	xs:string	Company internal short code for the current dossier. This value will be used as folder name for the regulatory action folder, see chapter 6.1. Example: “SYNCH127”
REFERENCED_DOSSIER	E	(0,*)	chapter 4.9.15	possibility to reference other dossiers In cases where the product has already been assessed by a regulator and a request is now being made a modification or renewal of an authorisation this element can be used to provide reference to the previous dossiers submitted to the regulator.
DOSSIER_RA	E	(1,*)	chapter 4.9.16	mandatory element to specify regulatory framework applicable for each receiver

SUBMISSION	E	(1,*)	chapter 4.9.17	Each GHSTS XML backbone file contains the SUBMISSION element for the current and all – if any – preceding submissions. Multiple instances are ordered by ascending submission number in the order of their appearance in the XML file.
------------	---	-------	----------------	--

Table 41 –child elements definition for element DOSSIER

4.9.15. REFERENCED_DOSSIER

It is possible to make a content reference into one or more other dossiers with the help of the element REFERENCED_DOSSIER_NUMBER, which was already submitted, for example

- under another "Regulatory Action" or
- from a Product dossier to an Active Ingredient dossier (in the pesticides domain).

4.9.15.1. Child elements

Element name	REFERENCED_DOSSIER			
Xpath	GHSTS/PRODUCT/DOSSIER/REFERENCED_DOSSIER			
Diagram				
Child element	T	Occ	Definition	Description
REFERENCED_DOSSIER_NUMBER	T	(1,1)	xs:string	The referenced dossier is identified with its PID (preferred) or a previous dossier numbering system for example the Caddy Dossier ID (e.g. LANDE005)
REFERENCED_DOSSIER_REASON	T	(1,1)	xs:string	human-readable reason why this reference is supplied in this submission

Table 42 –child elements definition for element REFERENCED_DOSSIER

4.9.16. DOSSIER_RA

For the explanation of the logical concept please see chapter 3.2.28.

4.9.16.1. Child elements

Element name	DOSSIER_RA			
Xpath	GHSTS/PRODUCT/DOSSIER/DOSSIER_RA			
Diagram				
Child element	T	Occ	Definition	Description
REGULATORY_TYPE	E	(1,1)	picklist, see chapter 3.2.24	Legislation, act, regulation or directive under which the submission is made. Examples: “91/414/EWG”, “1107/2009/EG,98/8/EG”
APPLICATION_TYPE	E	(1,1)	picklist, see chapter 3.2.24	Specifies the action type requested e.g. Joint Review, Authorisation. Modification, Re-evaluation, Renewal
PROJECT_ID_NUMBER	T	(0,*)	xs:string	The PROJECT_ID_NUMBER(s) is / are assigned by the Regulatory authority to the dossier. Example: “2011-1234”

Table 43 –child elements definition for element DOSSIER_RA

4.9.16.2.Attributes

Attribute name	Man	Type	Definition	Description
To_Specific_for_RA_Id	Y	XS:IDREF	chapter 4.5	ID reference to legal entity instance. Reference validated in schema using keyref RefIntegrityDossier_RAToLegalentity

Table 44 - attributes definition for element DOSSIER_RA

4.9.16.3.Constraints

Constraint Name	Description
	The element has to be supplied for all receiver of the submission

Table 45 - constraints definition for element DOSSIER_RA

4.9.17. SUBMISSION

For the explanation of the logical concept please see chapter 3.2.5.

4.9.17.1. Child elements

Element name	SUBMISSION			
Xpath	GHSTS/PRODUCT/DOSSIER/SUBMISSION			
Diagram				
Child element	T	Occ	Definition	Description
SUBMISSION_NUMBER	T	(1,1)	chapter 3.2.5	sequence of the submission represented as two digit number padded with zero, e.g. 01, 02, 03, ...
SUBMISSION_VERSION_DATE	T	(1,1)	xs:date	Date of the submission, in the form YYYY-MM-DD Example: "2009-04-07"
SUBMISSION_TITLE	T	(1,1)	xs:string	Title of the submission Example: "Spiroxamine OECD Annex II Update from March 2009"
INCREMENTAL	T	(1,1)	xs:boolean	Type of submission performed. For SUBMISSION_NUMBER = 1 INCREMENTAL=FALSE. For subsequent submissions the type of submission is dependent on the requirement specified by the regulator an INCREMENTAL=TRUE submission indicates only new or modified content is include in the submission package (see chapter 3.2.22)

Table 46 –child elements definition for element SUBMISSION

4.9.17.2. Constraints

Constraint Name	Description
	The first submission has have the submission number 01 and has to be of type INCREMENTAL=FALSE

Table 47 - constraints definition for element SUBMISSION

4.9.18. PRODUCT_USE

This element describes the use of the product, e.g. FEED, FOOD, GREENHOUSE, INDOOR, AGRICULTURE, AMENITY. The usage of the product defines the populations which may be exposed to the product and as a consequence the type of evidence (in the form of documents and data) required for the risk assessment.

4.9.19. PRODUCT_FUNCTION

Function of the product, eventually dependent on the chosen product use; e.g. Bactericide, Fungicide, Repellent, Novel Food.

4.9.20. COMPONENTS

This element represents a list of components (see chapter 3.2.10 for an explanation of the logical concept).

4.9.20.1. Child elements

Element name	COMPONENTS			
Xpath	GHSTS/COMPONENTS			
Diagram				
Child element	T	Occ	Definition	Description
COMPONENT	E	(1,*)	chapter 4.9.21	element for component

Table 48 –child elements definition for element COMPONENTS

4.9.21. COMPONENT

For the explanation of the logical concept please see chapter 3.2.10.

4.9.21.1. Child elements

Element name	COMPONENT			
Xpath	GHSTS/COMPONENTS/COMPONENT			
Diagram				
Child element	T	Occ	Definition	Description
METADATA_STATUS	E	(1,1)	picklist, see chapter 3.2.24	see chapter 3.2.3

COMPONENT_NAME	T	(1,1)	xs:string	common name of the component if available otherwise a registrant internal name Example: "Spiroxamine"
COMPONENT_PID	T	(0,1)	PIDType, chapter 5.2.9	PID of the pure reference component, see chapter 3.2.26.
COMPONENT_CLASS	E	(1,1)	core picklist, see chapter 3.2.24	Currently used for distinction between (genetically modified) organism and (chemical) substance. The selected value will further determine applicable domain-specific values for COMPONENT_IDENTIFIER and COMPONENT_DESCRIPTION.
COMPONENT_IDENTIFIER	E	(0,*)	chapter 4.6	Identity of the component The identifier type (registry/source) and the identifier must be provided Example of Chemical Abstracts Number COMPONENT_IDENTIFIER_TYPE.CODE = 'CASNO' COMPONENT_IDENTIFIER_TYPE.DECODE = 'Chemical Abstracts Service Number' COMPONENT_IDENTIFIER.VALUE = '118134-30-8' It is recommended to provide multiple identifiers to fully characterise a component.
COMPONENT_DESCRIPTION			chapter 4.6	One or more component descriptors to characterise the identity of the component The description type (e.g. species) and the description must be provided Example of Mild Pepino mosaic virus isolate VX1 COMPONENT_DESCRIPTION_TYPE.CODE = 'SPECIES' COMPONENT_DESCRIPTION_TYPE.DECODE = 'Species' COMPONENT_DESCRIPTION.VALUE = 'Mild Pepino mosaic virus' COMPONENT_DESCRIPTION_TYPE.CODE = 'STRAIN' COMPONENT_DESCRIPTION_TYPE.DECODE = 'Strain' COMPONENT_DESCRIPTION.VALUE = 'VX1' It is recommended to provide multiple descriptors to fully characterise a component.

Table 49 –child elements definition for element COMPONENT

4.9.21.2. Attributes

Attribute name	Man	Type	Definition	Description
Id	Y	XS:ID	chapter 4.5	unique ID, which is used for reference in other elements

Table 50 - attributes definition for element COMPONENT

4.9.21.3. Constraints

Constraint Name	Description
	For the pesticides domain: If a CAS-Number is supplied in COMPONENT_IDENTIFIER: Check the validity with CAS-Rules http://www.cas.org/content/chemical-substances/checkdig

Table 51 - constraints definition for element COMPONENT

4.9.22. DOCUMENTS

The element DOCUMENTS contains a list of documents.

4.9.22.1. Child elements

Element name	DOCUMENTS			
Xpath	GHSTS/DOCUMENTS			
Diagram	<pre> classDiagram class DOCUMENTS class DOCUMENT DOCUMENTS "1" -- "1..∞" DOCUMENT </pre>			
Child element	T	Occ	Definition	Description
DOCUMENT	E	(1,*)	chapter 4.9.23	element representing one single document concept which can be comprised of one to many files

Table 52 –child elements definition for element DOCUMENTS

4.9.23. DOCUMENT

The DOCUMENT metadata were grouped in a generic branch (DOCUMENT_GENERIC) common for all RA and a branch of the RA-specific document metadata (DOCUMENT_RA). Both branches have an own element METADATA_STATUS to characterize the status of the metadata independently.

For the explanation of the logical concept please see chapter 3.2.12.

4.9.23.1. Child elements

Element name	DOCUMENT			
Xpath	GHSTS/DOCUMENTS/DOCUMENT			
Diagram				
Child element	T	Occ	Definition	Description
DOCUMENT_RA	E	(1,*)	chapter	There is one element for each RA at which the document is targeted.
DOCUMENT_GENERIC	E	(1,1)	chapter	Bibliographic metadata and additional descriptors that are not dependent on RA.

Table 53 –child elements definition for element DOCUMENT

4.9.23.2. Attributes

Attribute name	Man	Type	Definition	Description
Id	Y	XS:ID	chapter 4.5	unique ID, which is used for reference in other elements, e.g. to link the document to one or many ToC nodes.

Table 54 - attributes definition for element DOCUMENT

4.9.23.3. Constraints

Constraint Name	Description
	Each DOCUMENT instance must reference at least one ToC node via its ID, using the element TOC2DOC (see chapter 4.10.1.2).

Table 55 - constraints definition for element DOCUMENT

4.9.24. DOCUMENT_RA

The element DOCUMENT_RA is present for each RA to which the registrant wants to assign the document. In addition the element contains RA specific information that may differ between the different RA receiving the same document. The information that is invariant for all RA is represented below the element DOCUMENT_GENERIC (see chapter 4.9.28).

For the explanation of the logical concept please see chapter 3.2.28.

4.9.24.1. Child elements

Element name	DOCUMENT_RA			
Xpath	GHSTS/DOCUMENTS/DOCUMENT/DOCUMENT_RA			
Diagram				
Child element	T	Occ	Definition	Description
METADATA_STATUS	E	(1,1)	picklist, see chapter 3.2.24	see chapter 3.2.3
DATA_PROTECTION_CLAIM	E	(1,1)	picklist, see chapter 3.2.24	indicates a data protection claim of the submitter Values are Y,N,O (O=Open)
DATA_REQUIREMENT	E	(1,1)	picklist, see chapter 3.2.24	Information about the necessity of the document for the RA Values are Y,N,O (O=Open) <ul style="list-style-type: none"> • Y: The document is mandatory according to the guidelines of the RA • N: The document is optional or not required according to the guidelines of the RA • O: The necessity is undefined
DOCUMENT_COMMENT	T	(0,*)	xs:string	regulatory authority specific document comments Example: “Study number: T0035534, Report number 20416”
OTHER_NATIONAL_GUIDELINE	E	(0,*)	chapter 4.9.25	additional Regulatory authority specific references into other national guidelines
RA_DOCUMENT_NUMBER	E	(0,1)	chapter 4.9.26	element to submit document identifiers already assigned by the RA

Table 56 –child elements definition for element DOCUMENT_RA

The correct usage and the correct interpretation of the elements DATA_PROTECTION_CLAIM and DATA_REQUIREMENT depend from the Regulatory authority and the specific regulatory action.

4.9.24.2.Attributes

Attribute name	Man	Type	Definition	Description
To_Specific_for_RA_Id	Y	XS:IDREF	chapter 4.5	ID reference to legal entity instance. Reference validated in schema using keyref RefIntegrityDocument_RAToLegalentity

Table 57 - attributes definition for element DOCUMENT_RA

4.9.24.3.Constraints

Constraint Name	Description
	The element has to be supplied for at least one receiver of the submission.
	For every RA referenced by the DOCUMENT_RA there must be at least one referenced FILE instance of this document with an element FILE_RA, where this same RA appears. (Rationale: A document without content for a RA does not make sense).
	Only those RA, that appear (as reference) in DOCUMENT_RA can be used in the FILE_RA elements of the referenced FILE instances of this document. (Rationale: The FILE_RA can only assign content to RA, to which the parent document is targeted.)

Table 58 - constraints definition for element DOCUMENT_RA

4.9.25. OTHER_NATIONAL_GUIDELINE

With the help of the non-mandatory element OTHER_NATIONAL_GUIDELINE and its child elements GUIDELINE_SYSTEM and GUIDELINE_NUMBER it is possible to make additional Regulatory authority specific references into other national guidelines.

4.9.25.1.Child elements

Element name	OTHER_NATIONAL_GUIDELINE			
Xpath	GHSTS/DOCUMENTS/DOCUMENT/ DOCUMENT_RA/OTHER_NATIONAL_GUIDELINE			
Diagram				
Child element	T	Occ	Definition	Description
GUIDELINE_SYSTE M	T	(1,1)	xs:string	guideline system e.g. DACO, OPPTS
GUIDELINE_NUMB ER	T	(1,1)	xs:string	identifier in the specified guideline system

Table 59 –child elements definition for element OTHER_NATIONAL_GUIDELINE

4.9.26. RA_DOCUMENT_NUMBER

4.9.26.1. Child elements

Element name	RA_DOCUMENT_NUMBER			
Xpath	GHSTS/DOCUMENTS/DOCUMENT/ DOCUMENT_RA/RA_DOCUMENT_NUMBER			
Diagram				
Child element	T	Occ	Definition	Description
RA_DOCUMENT_NUMBER_TYPE	E	(1,1)	chapter 4.6	
IDENTIFIER	E	(1,1)	xs:string	document identification number assigned by regulatory authority, e.g. the MRID. This element can also be used to identify documents that have been submitted before the usage of GHSTS.
ALREADY_SUBMITTED	T	(1,1)	xs:boolean	true, if this document has already been submitted in one or many different regulatory actions. The document PID needs to be maintained in these cases. Details can be provided in the element DOSSIER_CONTEXT.
DOSSIER_CONTEXT	E	(0,*)	chapter 4.9.27	element to provide details about previous submission context(s) of this document

Table 60 –child elements definition for element RA_DOCUMENT_NUMBER

4.9.27. DOSSIER_CONTEXT

This element allows the provision of information about a submission of a document in a different dossier, either by a dossier PID for GHSTS dossiers or by textual information for legacy dossiers of other formats, for example the legacy dossier number.

4.9.27.1. Child elements

Element name	DOSSIER_CONTEXT			
Xpath	GHSTS/DOCUMENTS/DOCUMENT/ DOCUMENT_RA/RA_DOCUMENT_NUMBER/DOSSIER_CONTEXT			
Diagram				
Child element	T	Occ	Definition	Description
DOSSIER_PID (choice)	T	(1,1)	PIDType, chapter 5.2.9	PID of the dossier (for GHSTS dossiers)
DOSSIER_NUMBER (choice)	T	(1,1)	xs:string	dossier number of the dossier (for legacy dossiers)

Table 61 –child elements definition for element DOSSIER_CONTEXT

4.9.28. DOCUMENT_GENERIC

This element groups the information about a document that is identical for all RA.

4.9.28.1. Child elements

Element name	DOCUMENT_GENERIC
Xpath	GHSTS/DOCUMENTS/DOCUMENT/DOCUMENT_GENERIC
Diagram	<p>The diagram illustrates the structure of the DOCUMENT_GENERIC element. It is a container element that includes the following child elements:</p> <ul style="list-style-type: none"> METADATA_STATUS DOCUMENT_PID DOCUMENT_COMPANY_NUMBER (dashed box) DOCUMENT_FAMILY_PID DOCUMENT_FAMILY DOCUMENT_TYPE (dashed box) DOCUMENT_CONTENT_STATUS_HISTORY REFERENCED_DOCUMENT (dashed box, cardinality 0..∞) RELATED_TO_COMPONENT (dashed box, cardinality 0..∞) DOCUMENT_NUMBER (dashed box, cardinality 0..∞) DOCUMENT_TITLE DOCUMENT_AUTHOR DOCUMENT_ISSUE_DATE DOCUMENT_OWNER (cardinality 1..∞) PUBLISHED_INDICATOR A choice element containing: <ul style="list-style-type: none"> COMPLETE_DOCUMENT_SOURCE Another choice element containing: <ul style="list-style-type: none"> DOCUMENT_YEAR (dashed box) DOCUMENT_ISSUE (dashed box) DOCUMENT_VOLUME (dashed box) DOCUMENT_PAGES (dashed box) TEST_LABORATORY (dashed box, cardinality 0..∞) GXP_INDICATOR TESTED_ON_VERTTEBRATE REFERENCED_TO_FILE (cardinality 1..∞)

Child element	T	Occ	Definition	Description
METADATA_STATUS	E	(1,1)	picklist, see chapter 3.2.24	see chapter 3.2.3.3
DOCUMENT_PID	T	(1,1)	PIDType, chapter 5.2.9	unique PID of the document, see chapter 3.2.26.
DOCUMENT_COMPANY_NUMBER	T	(0,1)	xs:string	Identifier of the document in the company internal dossier management system, e.g. for use in human communication. Or a DOI/URI if this document is publically available
DOCUMENT_FAMILY_PID	T	(1,1)	PIDType, chapter 5.2.9	A machine readable filter to select documents which belongs to one "document family", see chapter 3.2.13.
DOCUMENT_FAMILY	T	(1,1)	xs:string	A human readable description of the documents and files which together form a piece of evidence to be assessed under the regulatory action.
DOCUMENT_TYPE	T	(0,1)	picklist, see chapter 3.2.24	Categorisation of documents; in addition to the ToC node assignment (see chapter 3.2.15)
DOCUMENT_CONTENT_STATUS_HISTORY	E	(1,1)	picklist, see chapter 3.2.24	Used to indicate the status of the document (e.g. content modifications) in every submission, see chapter 3.2.18
REFERENCED_DOCUMENT	E	(0,*)	chapter 4.9.30	references to other documents. Can e.g. be used when a scientific summary makes references to study reports or scientific articles which are included in the dossier and make up the evidence base supporting the summary.
RELATED_TO_COMPONENT	E	(0,*)	chapter 4.9.31	references to components Where the safety of active substance and formulants have been assessed in separate studies, this allows those documents to be linked to the relevant substance from the COMPONENT complex element
DOCUMENT_NUMBER	E	(0,*)	picklist, see chapter 3.2.24	For usage see comment below this table.
DOCUMENT_TITLE	T	(1,1)	xs:string	Full title of the document provided by the registrant. This value will be used for reference lists.
DOCUMENT_AUTHOR	T	(1,1)	xs:string	For usage see comment below this table
DOCUMENT_ISSUE_DATE	T	(1,1)	xs:date	Date of the document
DOCUMENT_OWNER	T	(1,*)	xs:string	Owner of the document at the time of the current submission. If more than one owner exists, repeat the element.

PUBLISHED_INDICATOR	T	(1,1)	xs:boolean	Indicates whether a document has been published and therefore made publicly available. For usage see comment below this table
COMPLETE_DOCUMENT_SOURCE (choice option 1)	T	(1,1)	xs:string	Bibliographic citation source For usage see comment below this table
DOCUMENT_SOURCE (choice option 2)	T	(1,1)	xs:string	Bibliographic citation source For usage see comment below this table
DOCUMENT_YEAR (choice option 2)	T	(0,1)	xs:positiveInteger	For usage see comment below this table
DOCUMENT_ISSUE (choice option 2)	T	(0,1)	xs:string	For usage see comment below this table
DOCUMENT_VOLUME (choice option 2)	T	(0,1)	xs:string	For usage see comment below this table
DOCUMENT_PAGES (choice option 2)	T	(0,1)	xs:string	For usage see comment below this table
TEST_LABORATORY	T	(0,*)	xs:string	Name and address of the test facility / facilities where the study was carried out. Example: "Bayer AG, Wuppertal, Germany"
GPX_INDICATOR	T	(1,1)	xs:boolean	Indicates on the document level whether the test facility was carried out under accredited Good Practices. GXP was used as a general term for GLP, GAP, GBP.
TESTED_ON_VERTEBRATE	T	(1,1)	xs:boolean	Indicates whether a study has been conducted on a vertebrate
REFERENCED_TO_FILE	E	(1,*)	chapter 4.9.32	Connects the DOCUMENT level with the FILE level. It is possible to reference more than one FILE. The reference to the FILE instance with FILE_TYPE="Main" has to appear first in the XML file.

Table 62 –child elements definition for element DOCUMENT_GENERIC

DOCUMENT_NUMBER

The registrant often has a list of historical and additional numbers for each document. The reason for the existence of different numbers on the cover page can be multifaceted. For example, a report of a residue trial can be marked by the following numbers:

- Report number
- Study Code
- Project No.

- Plot No(s).
- Company DocID
- Old document number

Those kinds document number types can be submitted via the multi valued and picklist-controlled element DOCUMENT_NUMBER.

In addition, DOIs (Digital Object Identifier) can be supplied as document numbers. DOIs are used to uniquely identify digital objects and are widely used to identify academic, professional, and government information. The DOI is a link which resolves to the correct resource or service location.

As of now, a DOI does not replace the PID identifier nor releases the constraint that all content that is part of the submission must be physically present in the submission package.

As a DOCUMENT in GHSTS can be a compound of multiple FILES, but usually a value in DOCUMENT_NUMBER represents only one computer file, the DOCUMENT_NUMBER references the “Main” FILE as core representation of the DOCUMENT. The DOI mechanism supports referencing an object with multiple representations and could therefore mimic the GHSTS modelling.

DOCUMENT_AUTHOR

The element DOCUMENT_AUTHOR contains the author(s) of a document in one string, which is to be used unchanged in the content of document type “reference lists”. It should be possible for Regulatory authorities to split the authors’ string into individual authors while importing the document data. Therefore it is necessary to consider the following business rules:

Each Author is cited by Surname + comma + blank + each first initial of first name + point. In case of multiple first names use a blank between first names. Titles (e.g. Dr.) are not a part of the author name. Authors are separated by semicolon + blank. Don't use concatenation of authors with "and". If there are more than 3 authors, it is possible to end with "et al." without a semicolon.

Examples:

- Roemer, K. G.; Torres Alanis, O.; Garcia de Torres, G.
- Whitford, F. A.; Kronenberg, J.; Lunchick, C. et al.

PUBLISHED_INDICATOR

PUBLISHED_INDICATOR indicates whether a document has been published and therefore made publicly available.

- If the PUBLISHED_INDICATOR is false repeat the owner of the document in the element COMPLETE_DOCUMENT_SOURCE.
- If the document is published (PUBLISHED_INDICATOR = true) submit the publication information according to one of the following two choices:
- Use the OECD citation guidance.

Set the value of COMPLETE_DOCUMENT_SOURCE to one string with the full source citation like Journal, [issue], [volume], (year): pages.

- Split the bibliographic data in detail fields: DOCUMENT_SOURCE (contains the title of the journal / book), DOCUMENT_YEAR, DOCUMENT_ISSUE, DOCUMENT_VOLUME and DOCUMENT_PAGES.

It is recommended that reference lists contain as DOCUMENT_SOURCE the concatenation of all non-empty elements of the split bibliographic data detail fields listed above.

4.9.28.2. Constraints

Constraint Name	Description
	The element DOCUMENT_NUMBER is mandatory for unpublished documents (PUBLISHED_INDICATOR = false).
	The value of DOCUMENT_TITLE will be used unchanged in the content of reference lists (The check of this constraint is outside of the scope of GHSTS).
	The value of DOCUMENT_AUTHOR will be used unchanged in the contents of reference lists (The check of this constraint is outside of the scope of GHSTS).
	The value of DOCUMENT_ISSUE_DATE will be used unchanged in the contents reference lists (The check of this constraint is outside of the scope of GHSTS).
	If the PUBLISHED_INDICATOR = true then repeat the value of DOCUMENT_OWNER in the element COMPLETE_DOCUMENT_SOURCE.
	At least one element REFERENCED_TO_FILE must exist. At least one referenced FILE instance must have FILE_TYPE="Main".
	In the values of the element REFERENCED_TO_FILE the reference to the FILE instance with FILE_TYPE="Main" has to appear first in the XML file.

Table 63 - constraints definition for element DOCUMENT_GENERIC

4.9.29. DOCUMENT_CONTENT_STATUS_HISTORY

The element DOCUMENT_CONTENT_STATUS_HISTORY contains the full document content history list from the initial appearance of the document in the dossier to the current submission. Please note that this history is not RA-specific.

For the explanation of the logical concept please see chapter 3.2.18.

4.9.29.1. Child elements

Element name	DOCUMENT_CONTENT_STATUS_HISTORY			
Xpath	GHSTS/DOCUMENTS/DOCUMENT/DOCUMENT_GENERIC/DOCUMENT_CONTENT_STATUS_HISTORY			
Diagram				
Child element	T	Occ	Definition	Description
DOCUMENT_CONTENT_STATUS	E	(1,*)	picklist, see chapter 3.2.24	See chapter 3.2.18

Table 64 –child elements definition for element DOCUMENT_CONTENT_STATUS_HISTORY

4.9.29.2. Attributes

Attribute name	Man	Type	Definition	Description
submission_number	Y		chapter 4.10.2	submission number of the associated document content status entry Examples: "01", "02", "03"

Table 65 - attributes definition for element DOCUMENT_CONTENT_STATUS_HISTORY

4.9.29.3. Constraints

Constraint Name	Description
	Multiple entries for DOCUMENT_CONTENT_STATUS_HISTORY in the XML file are sorted with ascending submission number
	There must be only one entry with CONTENT_STATUS_TYPE = "New" as first entry. For exceptional cases it is permitted to re-introduce a DOCUMENT instance D[x] to a dossier that has been retired before. In this case the content lifecycle status is also "New".
	If the CONTENT_STATUS_TYPE is set to "Retired" for a DOCUMENT instance then this value is continued to be set to "Retired" for all subsequent submissions (not to "No Change"). Also, see the exceptional cases for re-introduced DOUCMENT instances above.
	The CONTENT_STATUS_HISTORY is complete if there exists a unique record for each submission_number between the submission number with CONTENT_STATUS_TYPE ="New" and the current submission number.

Table 66 - constraints definition for element DOCUMENT_CONTENT_STATUS_HISTORY

4.9.30. REFERENCED_DOCUMENT

Using the element REFERENCED_DOCUMENT the registrant can supply supporting references from the current document to other documents that are part of the current GHSTS submission package or outside.

For the explanation of the logical concept please see chapter 3.2.29.

4.9.30.1. Child elements

Element name	REFERENCED_DOCUMENT			
Xpath	GHSTS/DOCUMENTS/DOCUMENT/DOCUMENT_GENERIC/REFERENCED_DOCUMENT			
Diagram				
Child element	T	Occ	Definition	Description
REFERENCE_TYPE	E	(1,1)	picklist, see chapter 3.2.24	The element specifies the semantics of the reference to the target document. Example: “document to submitted document”
INTERNAL	E	(1,1)	xs:boolean	The element specifies where the target document is internal (within the current dossier and its submissions) or external (outside of current dossier). Note: Internal references to documents that are only part of preceding submissions can only occur in the rare case of referencing replaced documents, as the XML backbone is “complete” in all other cases. Hence, internal references will most likely reference documents in the current GHSTS XML backbone.
DOCUMENT_PID (choice option 1)	T	(1,1)	PIDType, chapter 5.2.9	PID of target document must be supplied if INTERNAL is set to “true”. If “false”, DOCUMENT_PID should be supplied, if document already possesses a PID that is known to the receivers.
DOCUMENT_NUMBER (choice option 2)	E	(1,1)	chapter 4.6	A document number has to be supplied in cases, where a PID cannot be supplied. For usage comments see chapter 4.9.28.1.

Table 67 –child elements definition for element REFERENCED_DOCUMENT

For the XSD choice element in REFERENCED_DOCUMENT one of the two elements DOCUMENT_PID or DOCUMENT_NUMBER must be supplied.

4.9.31. RELATED_TO_COMPONENT

The element RELATED_TO_COMPONENT is useful to characterize, that a study is related only to one of the active substances, a metabolite, a comparable active substance. or an organism Example: This is useful for residue trials, where only one of the active substances is analysed.

Every referred component has to be listed under COMPONENTS.

For the explanation of the logical concept please see chapter 3.2.29.

4.9.31.1. Child elements

Element name	RELATED_TO_COMPONENT			
Xpath	GHSTS/DOCUMENTS/DOCUMENT/DOCUMENT_GENERIC/RELATED_TO_COMPONENT			
Diagram				
Child element	T	Occ	Definition	Description
This element does not possess child elements, only an attribute				

Table 68 –child elements definition for element RELATED_TO_COMPONENT

4.9.31.2. Attributes

Attribute name	Man	Type	Definition	Description
To_Component_Id	Y	XS:IDREF	chapter 4.5	ID reference to Component. Reference validated in schema using keyref RefIntegrityDocumentToComponent

Table 69 - attributes definition for element RELATED_TO_COMPONENT

4.9.32. REFERENCED_TO_FILE

For the explanation of the logical concept please see chapter 3.2.12.

4.9.32.1. Child elements

Element name	REFERENCED_TO_FILE			
Xpath	GHSTS/DOCUMENTS/DOCUMENT/DOCUMENT_GENERIC/REFERENCED_TO_FILE			
Diagram				
Child element	T	Occ	Definition	Description
This element does not possess child elements, only an attribute				

Table 70 –child elements definition for element REFERENCED_TO_FILE

4.9.32.2. Attributes

Attribute name	Man	Type	Definition	Description
To_File_Id	Y	XS:IDREF	chapter 4.5	ID reference to FILE instance. Reference validated in schema using keyref RefIntegrityDocumentToFile

Table 71 - attributes definition for element REFERENCED_TO_FILE

4.9.33. FILES

The element FILES contains a list of files.

4.9.33.1. Child elements

Element name	FILES			
Xpath	GHSTS/FILES			
Diagram				
Child element	T	Occ	Definition	Description
FILE	E	(1,*)	chapter 4.9.34	single FILE instance

Table 72 –child elements definition for element FILES

4.9.33.2. Constraints

Constraint Name	Description
	For each DOCUMENT instance: The sequence of FILEs inside of the element FILES has to be the file order to read the DOCUMENT instance. This means that for each DOCUMENT instance the FILE instance with FILE_TYPE="Main" has to appear first in the XML file.

Table 73 - constraints definition for element FILES

4.9.34. FILE

For the explanation of the logical concept please see chapter 3.2.11.

4.9.34.1. Child elements

Element name	FILE			
Xpath	GHSTS/FILES/FILE			
Diagram				
Child element	T	Occ	Definition	Description
FILE_RA	E	(1,*)	chapter 4.9.35	element to specify at which RA the file is targeted and to supply additional metadata. Each file needs to be targeted to at least one receiving RA.
FILE_GENERIC	E	(1,1)	chapter 4.9.36	element with metadata not dependent on RA

Table 74 –child elements definition for element FILE

4.9.34.2.Attributes

Attribute name	Man	Type	Definition	Description
Id	Y	XS:ID	chapter 4.5	unique ID, which is used for reference in other elements

Table 75 - attributes definition for element FILE

4.9.35. FILE_RA

With the help of the element FILE_RA it is possible to send Regulatory authority specific metadata for each computer file.

The element FILE_RA is present for each RA to which the registrant wants to assign a specific file. In addition the element contains RA specific information that may differ between the different RA receiving the same file. The information that is invariant for all RA is represented below the element FILE_GENERIC (see chapter 4.9.36).

For the explanation of the logical concept please see chapter 3.2.28.

4.9.35.1.Child elements

Element name	FILE_RA			
Xpath	GHSTS/FILES/FILE/FILE_RA			
Diagram				
Child element	T	Occ	Definition	Description
METADATA_STATUS	E	(1,1)	picklist, see chapter 3.2.24	see chapter 3.2.3.3
CBI_DESIGNATION_CLAIM	T	(1,1)	picklist, see chapter 3.2.24	indicates whether the submitter claims for this RA that the file contains CBI-protected information (CBI = Confidential Business Information). Please note that the folder organisation of the computer files is independent of the value of this element; there is no physical separation into different subfolders.
FILE_COMMENT	T	(0,1)	xs:string	allows the registrant to communicate questions, notes, or other remarks to the agency at the FILE level

Table 76 –child elements definition for element FILE_RA

4.9.35.2.Attributes

Attribute name	Man	Type	Definition	Description
To_Specific_for_RA_Id	Y	XS:IDREF	chapter 4.5	ID reference to legal entity instance. Reference validated in schema using keyref RefIntegrityFILE_RAToLegalentity

Table 77 - attributes definition for element FILE_RA

4.9.35.3.Constraints

Constraint Name	Description
	The element has to be supplied for all RA to which the parent DOCUMENT instance is assigned according to the DOCUMENT_RA elements.

Table 78 - constraints definition for element FILE_RA

4.9.36. FILE_GENERIC

This element groups all metadata on the FILE level that is identical for all receiving RA.

4.9.36.1.Child elements

Element name	FILE			
Xpath	GHSTS/FILES/FILE_GENERIC			
Diagram	<pre> graph LR FG[FILE_GENERIC] --- MS[METADATA_STATUS] FG --- FPID[FILE_PID] FG --- FCI[FILE_COMPANY_ID] FG --- FT[FILE_TYPE] FG --- MT[MEDIA_TYPE] FG --- FC[FORMAT_COMMENT] FG --- MCS[MDSCHECKSUM] FG --- FN[FILENAME] </pre>			
Child element	T	Occ	Definition	Description
METADATA_STATUS	E	(1,1)	picklist, see chapter 3.2.24	see chapter 3.2.3
FILE_PID	T	(1,1)	chapter 5.2.9	see chapter 3.2.26
FILE_COMPANY_ID	T	(0,1)	xs:string	registrant internal file identification number Example: "123789456A1b", "MO-99-009136-001"
FILE_TYPE	E	(1,1)	picklist, see chapter 3.2.24	See chapter 3.2.12 Example: "Main"

MEDIA_TYPE	E	(0,1)	xs:string	Describes the media type (technical format) of the FILE. Official Internet Assigned Numbers Authority (IANA) definitions should be used (see https://www.iana.org/assignments/media-types/media-types.xhtml) Note: New media types can easily be registered with IANA.
FORMAT_COMMENT	T	(0,1)	xs:string	optional textual comment to explain the file format / media type and the originating applications including the application version, to assist the receiver in opening the file in the correct application, also when significant time between initial submission and retrieval has passed.
MD5CHECKSUM	T	(0,1)	chapter 5.2.10	Checksum of the computer file, generated by the registrant. Can be used by receiver to detect data transmission errors between sender – transport medium – receiver. Not relevant for statement about content changes. Example: “79054025255fb1a26e4bc422aef54eb4”
FILENAME	T	(1,1)	chapter 5.2.7	see chapter 6.4

Table 79 –child elements definition for element FILE_GENERIC

4.9.36.2. Constraints

Constraint Name	Description
	A value for MD5CHECKSUM is mandatory if a computer file is being physically submitted in a dossier for the first time as part of a FILE instance with a specific PID. For subsequent submissions within the same regulatory action, the MD5CHECKSUM is mandatory, when the PID of the FILE instance changes (meaning that there is a semantic content change). For technical changes the MD5CHECKSUM information is optional.

Table 80 - constraints definition for element FILE_GENERIC

4.9.37. TOC

The ToC branch in the GHSTS XML backbone contains the hierarchical Table of Contents. The information is a near 1:1 copy of an instance of an officially supplied standard ToC XML file (as additional non-core GHSTS component, see chapter 2.4.2.5), enriched with a few additional attributes. It is not permitted to alter element text values for all elements that are direct copies from the standard ToC XML file. For these elements the description below will contain the information “as supplied with standard

In case there is no officially published ToC template available, the submitter has to generate the required ToC structure itself; according to the regulations and application type.

For the explanation of the logical concept please see chapter 3.2.7.

4.9.37.1. Child elements

Element name	TOC			
Xpath	GHSTS/TOC			
Diagram				
Child element	T	Occ	Definition	Description
METADATA_STATUS	E	(1,1)	picklist, see chapter 3.2.24	see chapter 3.2.3
TOC_FULL_NAME	T	(1,1)	xs:string	as supplied with standard ToC, created by submitter in other cases
TOC_SHORT_NAME	T	(1,1)	xs:string	dto.
TOC_PID	T	(0,1)	Pid Type	as supplied with standard ToC, empty in other cases
TOC_VERSION_TITLE	T	(1,1)	xs:string	as supplied with standard ToC, created by submitter in other cases
TOC_VERSION	T	(1,1)	versionnumberType	as supplied with standard ToC, created by submitter in other cases
TOC_OWNER	E	(1,1)	xs:string	as supplied with standard ToC, created by submitter in other cases
STANDARD_TOC_REFERENCE	E	(1,1)	chapter 4.9.38	see chapter 3.2.7
STRUCTURE	E	(0,1)	chapter 4.9.39	as supplied with standard ToC, created by submitter in other cases

Table 81 –child elements definition for element TOC

4.9.38. STANDARD_TOC_REFERENCE

STANDARD_TOC_REFERENCE provides information on what standard ToC is used in the GHSTS XML backbone. The GHSTS validator will use this information to validate the embedded ToC below the TOC element against the standard ToC.

4.9.38.1. Child elements

Element name	STANDARD_TOC_REFERENCE			
Xpath	GHSTS/TOC/STANDARD_TOC_REFERENCE			
Diagram				
Child element	T	Occ	Definition	Description
STANDARD_TOC_PID	T	(1,1)	PIDType, chapter 5.2.9	PID of the standard ToC, see chapter 3.2.26.
FILENAME	T	(1,1)	filenamestringType, chapter 5.2.7	relative path to Toc file in Toc folder, see chapter 6.1.

Table 82 –child elements definition for element STANDARD_TOC_REFERENCE

4.9.38.2. Constraints

Constraint Name	Description
	If a standard ToC is referenced, validate the TOC_NODES / TOC_PID below the element TOC in the GHSTS backbone XML file against the standard ToC.

Table 83 - constraints definition for element STANDARD_TOC_REFERENCE

4.9.39. STRUCTURE

STRUCTURE is at the top level of the ToC hierarchy.

4.9.39.1. Child elements

Element name	STRUCTURE			
Xpath	GHSTS/TOC/STRUCTURE			
Diagram				
Child element	T	Occ	Definition	Description
TOC_NODE	E	(1,*)	ToC node type, see chapter 4.10.1	

Table 84 –child elements definition for element STRUCTURE

4.9.40. LEGAL_ENTITIES

The element LEGAL_ENTITIES is a list of all Legal Entities which are used as a RECEIVER (Regulatory authorities) or as a SENDER (registrant)

4.9.40.1. Child elements

Element name	LEGAL_ENTITIES			
Xpath	GHSTS/LEGAL_ENTITIES			
Diagram				
Child element	T	Occ	Definition	Description
LEGAL_ENTITY	E	(1,*)	chapter 4.9.41	element for legal entity

Table 85 –child elements definition for element LEGAL_ENTITIES

4.9.41. LEGAL_ENTITY

The element LEGAL_ENTITY contains the metadata of one Legal Entity.

For the explanation of the logical concept please see chapter 3.2.8.

4.9.41.1. Child elements

Element name	LEGAL_ENTITY			
Xpath	GHSTS/LEGAL_ENTITIES/LEGAL_ENTITY			
Diagram				
Child element	T	Occ	Definition	Description
METADATA_STATUS	E	(1,1)	picklist, see chapter 3.2.24	see chapter 3.2.3
LEGALENTITY_PID	T	(1,1)	PIDType, chapter 5.2.9	unique PID of the Legal Entity, see chapter 3.2.26. Use of the “nil PID” is permitted.

LEGAENTITY_NAME	T	(1,1)	xs:string with attribute xml:space="preserve"	standard name of the Legal Entity. Line feeds are allowed in this element.
LEGAENTITY_TYPE	E	(0,1)	picklist, see chapter 3.2.24	e.g. "Company", "Regulatory Authority"
OTHER_NAME	T	(0,*)	xs:string	alternative name of the Legal Entity.
LEGAENTITY_IDENTIFIER	E	(0,*)	chapter 4.6	Characterize Legal Entities by other worldwide established identifier e.g. the DUNS-number or the VAT-number.
CONTACT_ADDRESS	E	(0,1)	chapter 4.9.42	element for the official contact of the legal entity
CONTACT_PERSON	E	(0,*)	chapter 4.9.43	element for content person(s)

Table 86 –child elements definition for element LEGAL_ENTITY

A replacement of a LEGAENTITY_NAME by a newly assigned common name will not impact the previously assigned LEGAENTITY_PID.

4.9.41.2. Attributes

Attribute name	Man	Type	Definition	Description
Id	Y	XS:ID	chapter 4.5	unique ID, which is used for reference in other elements

Table 87 - attributes definition for element LEGAL_ENTITY

4.9.42. CONTACT_ADDRESS

Each Legal Entity can optionally be characterized by a CONTACT_ADDRESS.

4.9.42.1. Child elements

Element name	CONTACT_ADDRESS			
Xpath	GHSTS/LEGAL_ENTITIES/LEGAL_ENTITY/CONTACT_ADDRESS			
Diagram	<p>The diagram illustrates the structure of the CONTACT_ADDRESS element. It is represented as a dashed box on the left, connected by a line to a central connector box. From this connector, a vertical dashed line leads to a series of child elements, each in a dashed box: STREET1, STREET2, ZIPCODE, CITY, STATE, COUNTRY, PHONE, FAX, EMAIL, and WEBSITE. The COUNTRY element has a small square icon with a plus sign next to it, indicating it is a choice or list element.</p>			
Child element	T	Occ	Definition	Description
STREET1	T	(0,1)	xs:string	Part of postal address Examples: “Geschäftsbereich Pflanzenschutz”, “608E-2720 Riverside Drive“
STREET2	T	(0,1)	xs:string	Part of postal address Examples: “Alfred-Nobel-Straße 50”, “Tupper Building, E6607”
ZIPCODE	T	(0,1)	xs:string	codes established by postal authorities for purposes of sorting and delivering mail. Example: “D-40789”
CITY	T	(0,1)	xs:string	Represents the city, town, village, or hamlet. Named ‘town’ or ‘municipality’ in the Rose Model Example: “Monheim am Rhein”
STATE	T	(0,1)	xs:string	Represents the State, Province, and/or County.
COUNTRY	E	(0,1)	picklist, see chapter 3.2.24	
PHONE	T	(0,1)	xs:string	Telephone number including country code Example: “+01 613-736-3666”
FAX	T	(0,1)	xs:string	Fax number including country code Example: “+49 2173 38 1234”

EMAIL	T	(0,1)	xs:string	Email address of the official contact. Format will not be validated
WEBSITE	T	(0,1)	xs:anyURI	Website URL. Format will not be validated. Example: www.bayercropscience.com

Table 88 –child elements definition for element CONTACT_ADDRESS

4.9.42.2. Constraints

Constraint Name	Description
	If the optional element CONTACT_ADDRESS is supplied, the registrant has to take care that a reasonable subset of the elements are supplied. No validation will be done on the elements. The combination of LEGALENTITY_NAME, STREET1, STREET2, ZIPCODE, CITY, STATE and COUNTRY should be usable for postal service.

Table 89 - constraints definition for element CONTACT_ADDRESS

4.9.43. CONTACT_PERSON

It is possible to define one or more contact persons.

4.9.43.1. Child elements

Element name	CONTACT_PERSON			
Xpath	GHSTS/LEGAL_ENTITIES/LEGAL_ENTITY/CONTACT_PERSON			
Diagram				
Child element	T	Occ	Definition	Description
ORGANISATION	T	(0,1)	xs:string with attribute xml:space="preserve"	company name of the contact person if it is different from the parent LEGALENTITY_NAME.
DEPARTMENT	T	(0,1)	xs:string with attribute xml:space="preserve"	the department of the contact person

TITLE	T	(0,1)	xs:string	title / salutation of the individual
FIRSTNAME	T	(0,1)	xs:string	firstname of the person
LASTNAME	T	(0,1)	xs:string	surname of the person
PHONE	T	(0,1)	xs:string	Telephone number including country code Example: "+01 613-736-3666"
MOBILE	T	(0,1)	xs:string	Mobile number including country code Example: "+49 172 2345 678"
FAX	T	(0,1)	xs:string	Fax number including country code Example: "+49 2173 38 1234"
EMAIL	T	(0,1)	xs:string	Email address of the contact person. Format will not be validated

Table 90 –child elements definition for element CONTACT_PERSON

4.9.43.2. Constraints

Constraint Name	Description
	If the optional element CONTACT_PERSON is supplied the registrant has to take care that a reasonable subset of the elements are supplied. No validation will be done on the elements.

Table 91 - constraints definition for element CONTACT_PERSON

4.10. Basic XML types

The following basic XML types are defined in the GHSTS schema definition.

4.10.1. TOC node type

The ToC node type in a GHSTS XML submission is a near 1:1 copy of the ToC node type of a standard ToC XML file (see chapter 2.4.2.5), enriched with a few additional attributes.

It is not permitted to alter element text values for all elements that are direct copies from the standard ToC XML file. For these elements the description below will contain the information "as supplied with standard ToC template".

4.10.1.1. Child elements

Type name	TYPE_TOC_NODE			
Diagram	<p>The diagram illustrates the structure of the TYPE_TOC_NODE element. It is a container for several child elements: NODE_NAME, NODE_HEADING, TOC_NODE_PID, EMPTY_NODE, TOC2DOC (0..∞), and TOC_NODE (0..∞). The TOC_NODE_PID element is shown with a dashed border, indicating it is optional. The TOC2DOC and TOC_NODE elements are shown with dashed borders and a plus sign icon, indicating they are optional and can occur multiple times (0..∞).</p>			
Child element	T	Occ	Definition	Description
NODE_NAME	T	(1,1)	xs:string	as supplied with standard ToC template
NODE_HEADING	T	(1,1)	xs:string	as supplied with standard ToC template
TOC_NODE_PID	T	(0,1)	PIDType, chapter 5.2.9	as supplied with standard ToC template, see chapter 3.2.26.
EMPTY_NODE	T	(1,1)	xs:boolean	Business information supplied by the registrant. If “true” this indicates that this node and all direct and indirect subnodes were deliberately left empty, if “false”, then at least one document is in current node or any subnode. Please note that this is not technical information, for example, a viewer program or ingestion component should rely on.
TOC2DOC	E	(0,*)		References from ToC node to document instances The sequential order of references to documents in the XML file is identical to the order in which the documents should appear in the ToC node. The order is relevant for referencing documents in reference lists.
TOC_NODE	E	(0,*)	TYPE_TOC_NODE	as supplied with standard ToC template

Table 92 –child elements definition for element TOC_NODE

4.10.1.2.TOC2DOC

Child elements

Element name	TOC2DOC			
Xpath	/TOC_NODE/TOC2DOC			
Diagram				
Child element	T	Occ	Definition	Description
NODE_ASSIGNMENT_STATUS	T	(1,1)	picklist, see chapter 3.2.24	Assignment status of the DOCUMENT I the ToC node, see chapter 3.2.20
REPLACED_DOCUMENT_PID	E	(0,1)	PIDType, chapter 5.2.9	When the node assignment status is set to “Replaced” then the PID of the replaced document from the same document family can optionally be provided.

Table 93 –child elements definition for element TOC2DOC

Attributes

Attribute name	Man	Type	Definition	Description
To_Document_Id				ID reference to document instance. Reference validated in schema using keyref RefIntegrityTocnodeToDocument

Table 94 - attributes definition for element TOC2DOC

4.10.2. Submission number type

The submission number consists of two digits, starting with 01, 02, 03, The submission number has to be used as the name of the submission folder for the submission package (see chapter 6.1).

4.10.2.1.Definition

```

<xs:simpleType name="submissionnumberType">
  <xs:annotation>
    <xs:documentation>definition of submission number as two digit number, e.g. 01,02,03 ...
  </xs:documentation>
</xs:annotation>
  <xs:restriction base="xs:string">
    <xs:pattern value="\d{2}"/>
  </xs:restriction>
</xs:simpleType>

```

4.11. Usage of external type definitions

The GHSTS XSD uses XML type definitions from the CommonComponents XSD (see chapter 5.2).

Chapter 5. .Referenced schema definitions

5.1. Standard ToC schema definition

5.1.1. Overview

This chapter provides a technical reference for the Table of Contents schema definition. This chapter contains the most relevant aspects of the XSD, especially an explanation on how elements should be used in the business context. The complete definition is contained in the XSD itself.

5.1.2. Relation to the submission schema definition

The ToC schema definition is a GHSTS core component and serves to define ToC templates for specific regulations, which are published for reuse (see chapter 2.4.2.5). When using a ToC template in a GHSTS submission it is included into the submission package and referenced by the XML element STANDARD_TOC_REFERENCE (see chapter 4.9.38).

5.1.3. Filename and version

The ToC XSD is versioned using a version numbering scheme of type XX.XX.XX. It is composed of a two digit major, a two digit minor and a two digit patch level version number separated by single dots. Major versions are used for global functional changes, minor versions for local changes and patch level versions for error corrections. Patch level versions do not require new versions of other GHSTS core components (see chapter 2.4).

This version is contained in the version attribute of the root schema element in the XSD file.

The filename uses the same versioning scheme with the hyphen instead of the dot: The file naming convention of the ToC XSD is therefore

toc_XX-XX-XX.xsd

for example

toc_01-00-00.xsd

Please note that the filename does not contain the string “ghsts” as the ToC definitions next to their usage in the context of GHSTS can be used in other processes.

For the instances of the GHSTS ToC XSD the following naming convention is proposed

toc-<TOC_OWNER>-<TOC_SHORT_NAME>_XX-XX-XX.xml

for example

toc-occd-standard_01-00-00.xml

The values for ToC Owner and ToC short name are taken from the corresponding elements in the XML file (see below). Spaces or hyphens in the values are replaced with underscore.

This chapter describes the version 02.00.00 of the ToC XML schema, published on <http://www.oecd.org/chemicalsafety/submission-transport-standard/>.

5.1.4. Namespace

The ToC schema description file uses the following namespace definitions:

- `xmlns:xs="http://www.w3.org/2001/XMLSchema"`
- `xmlns="http://www.oecd.org/TOCS"`
- `targetNamespace="http://www.oecd.org/TOCS"`

5.1.5. Unique keys

The elements `TOC_PID` and `TOC_NODE_PID` are part of the identifiers that are globally managed (see chapter 3.2.26). Their values have to be globally unique.

The value of `TOC_NODE_NAME` has to be unique within one instance of a standard ToC.

5.1.6. Notation conventions

For notation conventions see chapter 4.7 and chapter 4.8.

5.1.7. Elements

5.1.7.1. Overview

An overview diagram of the complete XSD is provided as separate image file in PNG format, with the filename:

`toc_XX-XX-XX.png`.

5.1.7.2. TOC

TOC is the definition of the ToC for the standard ToC templates.

Child elements

Element name	TOC			
Xpath	TOC			
Diagram				
Child element	T	Occ	Definition	Description
TOC_FULL_NAME	T	(1,1)	xs:string	Name of ToC family to be displayed in viewer Example: “OECD Plant Protection Products ToC”
TOC_SHORT_NAME	T	(1,1)	xs:string	short internal identifier of the ToC family Example: “OECD PPP ToC”
TOC_PID	T	(1,1)	PIDType, chapter 5.2.9	unique PID of the ToC version, see chapter 3.2.26.
TOC_VERSION_TITLE	T	(1,1)	xs:string	Short identifier of the ToC version, e.g. “OECD Pesticide ToC 2005”
TOC_VERSION	T	(1,1)	versionnumberType	version of the ToC, e.g. “01.01.00”
TOC_OWNER	E	(1,1)	xs:string	Owner / managing entity of the ToC template
TO_USE_FROM	T	(1,1)	xs:date	ToC valid from this date
TO_USE_UNTIL	T	(0,1)	xs:date	ToC valid until this date
STRUCTURE	E	(1,1)	chapter 4.9.39	Root element of the hierarchical ToC structure. Same element definition as in the GHSTS XSD.

Table 95 –child elements definition for element TOC

5.1.7.3.STRUCTURE

STRUCTURE is the top-level of the ToC hierarchy.

Child elements

Element name	STRUCTURE			
Xpath	TOC/STRUCTURE			
Diagram				
Child element	T	Occ	Definition	Description
TOC_NODE	E	(1,*)	TOC node template type , see chapter 5.1.8.1	The top-down sequence of the TOC_NODE instances in the XML file reflects the order of the ToC nodes in the ToC.

Table 96 –child elements definition for element STRUCTURE

5.1.8. Basic types

The standard ToC schema definition contains the following type definitions

5.1.8.1.Template ToC node type

The ToC node type in a ToC template is slightly different from the ToC node type as defined for the GHSTS submission (see chapter 4.10.1)

Child elements

Type name	TYPE_TEMPLATE_TOC_NODE			
Diagram				
Child element	T	Occ	Definition	Description
NODE_NAME	T	(1,1)	xs:string	code of an annex point, unique within the complete ToC Examples: "0.1.6003", "IIA-5.2.1"

NODE_HEADING	T	(1,1)	xs:string	textual context as a complete heading for outputs Example: "Acute oral toxicity"
DOCUMENT_REFERENCES	T	(1,1)	xs:Boolean	True, if this node can contain document references of type TOC2DOC (see chapter 4.10.1.2) when used in GHSTS XML backbone. If false, no document references are permitted from/to this node.
TOC_NODE_PID	T	(1,1)	PIDType, chapter 5.2.9	unique PID of the ToC node, see chapter 3.2.26.
TOC_NODE	E	(0,*)	TYPE_TEMPLATE_TOC_NODE	The recursive top-down sequence of the TOC_NODE instances in the XML file reflects the order of the ToC nodes in the ToC

Table 97 –child elements definition for element TOC_NODE

Constraints

Constraint Name	Description
	The value for NODE_NAME has to be unique within the complete ToC
	The top-down sequence of the TOC_NODE instances in the XML file reflects the order of the ToC nodes in the ToC

Table 98 - constraints definition for element TOC_NODE

5.1.9. Usage of external type definitions

The ToC XSD uses XML type definitions from the CommonComponents XSD (see chapter 5.2), that it includes.

5.2. CommonComponents XSD

5.2.1. Overview

The CommonComponents XSD (or short CC XSD) is a library XSD and contains XML type definitions that are use across multiple GHSTS XSD. The CC XSD is always used in conjunction with another XSD, never on its own.

5.2.2. Usage in main XSD

The CommonComponents XSD is used by importing into the main XSD (e.g. submission XSD, ToC XSD) by the xs:import directive, e.g.

```
<xs:import namespace="http://www.oecd.org/GHSTS/CommonComponents"
schemaLocation="CommonComponents_02-00-00.xsd"/>
```

and defining the namespace “cc” in the schema declaration of the importing XSD, e.g

```
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns="http://www.oecd.org/GHSTS"
xmlns:cc="http://www.oecd.org/GHSTS/CommonComponents" targetNamespace="http://www.oecd.org/GHSTS"
elementFormDefault="qualified" attributeFormDefault="unqualified" version="02.00.00">
```

5.2.3. Filename and version

The CC XSD is versioned using a version numbering scheme of type XX.XX.XX. It is composed of a two digit major, a two digit minor and a two digit patch level version number separated by single dots. This version is contained in the version attribute of the root schema element in the XSD file.

The filename uses the same versioning scheme with the hyphen instead of the dot: The file naming convention of the CC XSD is therefore

CommonComponents_XX-XX-XX.xsd

for example

CommonComponents_02-00-00.xsd

This chapter describes the version 02.00.00 of the CC XML schema, published on <http://www.oecd.org/chemicalsafety/submission-transport-standard/>.

5.2.4. Namespace

The ToC schema description file uses the following namespace definitions:

- xmlns:xs="http://www.w3.org/2001/XMLSchema"
- xmlns="http://www.oecd.org/GHSTS/CommonComponents"
- targetNamespace="<http://www.oecd.org/GHSTS/CommonComponents>"

5.2.5. Picklist Item type

The picklist item type represents an instance of a term from the picklist model. This includes its identification (CODE) and the complementary DECODE value for human-readable display. The same CODE may have decode values in different languages, the “lang_code” attribute should be used to indicate when the language is not English (en).

A picklist item instance is always used in conjunction with an attribute “catalogue” with a predefined and fixed value for each picklist-controlled GHSTS element. It serves to assist a GHSTS builder (or a ToC builder) to select the correct picklist from the set of picklist values, as it uniquely references one picklist in the picklist model. The fixed catalogue names for core-picklists are defined in chapter 3.2.24.1.

For an example of picklist usage see **Figure 12 – PicklistItem Type** .

5.2.5.1. Child elements

Type name	PicklistItemType			
Diagram				
Child element	T	Occ	Definition	Description
CODE	T	(0,1)	xs:string	machine-readable CODE supplied by the governing entity of the respective picklist.
DECODE	T	(1,1)	xs:string	human-readable value to be visible in a GHSTS viewer. The sender has to select one decode value in a specific language, preferably in the language accepted by the recipient

Table 99 –child elements definition for type PicklistItemType

5.2.5.2. DECODE

Child elements

Element name	DECODE			
Diagram				
Child element	T	Occ	Definition	Description
n.a.				

Table 100 –child elements definition for element DECODE

Attributes

Attribute name	Man	Type	Definition	Description
lang_code	N	xs:string	ISO 639-1:2002	ISO 639-1:2002 two-letter code of the language. There are no decode values itself for the language code, a GHSTS viewer may implement to supply decode values for the language as part of the custom viewer localization process

Table 101 - attributes definition for element DECODE

5.2.6. String definition

There is currently no need for a customized string definition and all standard text elements are defined using the XML default string type `xs:string`. Therefore, the CommonComponents XSD does not contain a customized string definition type. Additional constraints (e.g. length) have to be imposed using validation rules.

Please note that there is no explicit directive for whitespace handling, meaning that the default value “preserve” is active. For the tree elements

- LEGALENTITY_NAME
- ORGANISATION
- DEPARTMENT

an additional directive “`xml:space=`”preserve” for the XML parser is an attribute.

5.2.7. Relative file path type

The type `filenamestringType` defines what a relative path from the GHSTS XML backbone to a computer file location on the file system should be like:

- References to computer files only use relative paths.
- The GHSTS backbone file is located in the submission folder (see chapter 6.1). The relative path always first moves one folder level up to the regulatory action folder using “..” and then descends into the submission folder, consisting of two digits (see chapter 4.10.2). This is true even if the relative path points to a computer file in the current submission.
- The path separator is always “/” independent of the current platform.
- The folder names and file names only support a restricted set of characters.
- The total length of the relative path is limited to 230 characters. This is due to the limitation of some file systems with a total path length of 255 characters.

The type `filenamestringType` is used in the following elements:

- GHSTS/FILES/FILE/FILENAME
- GHSTS/TOC/STANDARD_TOC_REFERENCE/FILENAME

For recommendations on how to choose good folder and file names see chapters 6.3 and 6.4.

5.2.7.1. Definition

```
<xs:simpleType name="filenamestringType">
  <xs:annotation>
    <xs:documentation>definition of relative paths for filenames, used in
FILE</xs:documentation>
  </xs:annotation>
  <xs:restriction base="xs:string">
    <xs:whiteSpace value="preserve"/>
    <xs:minLength value="1"/>
    <xs:maxLength value="230"/>
    <xs:pattern value="(\.\.\/)\d{2}/((([a-zA-Z0-9_\-\.]+)/*)([a-zA-Z0-9_\-\.\/()]+)"/>
  </xs:restriction>
```



```
</xs:simpleType>
```

5.2.7.2.Examples

```
<FILENAME>../02/attachments/C6313_0016-supp1.bmp</FILENAME>
```

5.2.8. Version number type

The version number type defines the versioning scheme used to specify the GHSTS specification to which a GHSTS submission package adheres.

5.2.8.1.Definition

```
<xs:simpleType name="versionnumberType">
  <xs:annotation>
    <xs:documentation>Version number: six digits, e.g. 01.00.00</xs:documentation>
  </xs:annotation>
  <xs:restriction base="xs:string">
    <xs:pattern value="\d{2}.\d{2}.\d{2}"/>
  </xs:restriction>
</xs:simpleType>
```

5.2.8.2.Example

```
<GHSTS specificationversion="01.00.00" xsi:schemaLocation="http://www.oecd.org/GHSTS GHSTS_V010000.xsd"
xmlns="http://www.oecd.org/GHSTS" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
```

5.2.9. PID type

The PID type describes how a Persistent Identifier for GHSTS is represented.

The PID is built according to the Uniform Resource Name (URN) definition, see e.g. http://de.wikipedia.org/wiki/Uniform_Resource_Name.

A PID in the form of the URN is built according to the pattern

```
urn:<namespace>:<value>
```

The namespace is an abbreviation for the body that has issued the URN. There is a list of officially registered URN namespaces (e.g. see <http://www.iana.org/assignments/urn-namespaces/urn-namespaces.xml>), however it is possible to use the URN schema without an officially registered namespace. For all PID generated for GHSTS, the namespace “ghsts” will be used, irrespective of the generating entity.

The value part has to be unique per namespace to make the complete PID unique.

For example, the existing IUCLID “UUID” examples (please note that IUCLID calls these identifiers UUIDs whereas they are not compliant with the standard UUID definition)

```
ECB5-d966af20-7dea-42ad-a7be-87278757dbab
IUC5-1a16228d-42a1-4ed3-a405-be8728e17f8e
```

can be directly expressed as PID for GHSTS:

```
urn:ghsts:d966af20-7dea-42ad-a7be-87278757dbab
urn:ghsts:1a16228d-42a1-4ed3-a405-be8728e17f8e
```

For the value part of the PID the UUID format is used. UUID stands for Universally Unique Identifier and is usually used in software development. However, as UUIDs can be easily created and all common

programming environments support the creation of UUIDs it can also ideally serve as format for the value part of the PID.

For mandatory PID that are by default supplied externally (e.g. for a ToC node) but where a PID is not available (e.g. an authority requested custom / *ad hoc* node created in a ToC), the registrant can temporarily use the so-called “nil PID” or “fake PID”.

urn:ghsts:00000000-0000-0000-0000-000000000000

in the submission instead of a real PID.

For details about the UUID format please see chapter 5.2.9.4.

5.2.9.1. Definition

```
<xs:simpleType name="PIDType">
  <xs:annotation>
    <xs:documentation>definition of Persistent Identifier (PID) according to
URN pattern as defined in RFC 2141</xs:documentation>
  </xs:annotation>
  <xs:restriction base="xs:anyURI">
    <xs:pattern value="urn:ghsts:([a-zA-Z0-9()+,.:=@;$!*'-]|%[0-9A-Fa-
f]{2})+"/>
  </xs:restriction>
</xs:simpleType>
```

5.2.9.2. Constraints

Constraint Name	Description
	The GHSTS validator will check that the PID format is used for a defined set of namespaces.

Table 102 - constraints definition for element PID

5.2.9.3. Example

```
<COMPONENT_PID>urn:ghsts:1a16228d-42a1-4ed3-a405-be8728e17f8e</COMPONENT_PID>
<DOCUMENT_PID>urn:ghsts:d966af20-7dea-42ad-a7be-87278757dbab </DOCUMENT_PID>
```

5.2.9.4. UUID definition

The UUID definition is provided here for the registrants that need to create PID with the UUID as value part.

Technically a UUID is a 16-byte (128-bit) number. It is represented in a canonical form by 32 hexadecimal digits, displayed in 5 groups separated by hyphens, in the form 8-4-4-4-12 for a total of 36 characters (32 digits and 4 hyphens). An example for a UUID in the representation is as follows:

761e6400-e26b-56d5-a482-663375330000

There are multiple subtypes of UUID, see http://en.wikipedia.org/wiki/Universally_unique_identifier. For UUIDs generated by the registrant (see “Table 12 – Overview of concepts identified by PID”), GHSTS doesn’t restrict the UUID version to be used but recommends to use one out of version 3 (MD5 hash), version 4 (random) and version 5 (SHA-1 hash).

All common programming languages support the creation of UUIDs. For details please refer to the programming language reference.

5.2.9.5. Definition

```
<xs:simpleType name="uuidType">
  <xs:annotation>
    <xs:documentation>definition of UUID according to canonical
representation</xs:documentation>
  </xs:annotation>
  <xs:restriction base="xs:string">
    <xs:pattern value="[a-f0-9]{8}\-[a-f0-9]{4}\-[a-f0-9]{4}\-[a-f0-9]{4}\-[a-f0-9]{12}" />
  </xs:restriction>
</xs:simpleType>
```

5.2.10. MD5 type

The MD5 type is used in GHSTS for the element MD5CHECKSUM of the FILE concept. The MD5 Message-Digest Algorithm is a widely used cryptographic hash function that produces a 128-bit (16-byte) hash value, represented as hexadecimal number of 32 digits. It is specified in RFC 1321. For details see <http://en.wikipedia.org/wiki/MD5>.

The value is supplied by the registrant to enable the receiver to check the successful transmission process of the referenced computer file.

All common programming languages support the creation of MD5 value for a given file. For details please refer to the programming language reference.

5.2.10.1. Definition














```
<xs:simpleType name="md5Type">
  <xs:annotation>
    <xs:documentation>MD5 checksum definition, used in FILE</xs:documentation>
  </xs:annotation>
  <xs:restriction base="xs:string">
    <xs:pattern value="[a-f,A-F,0-9]{32}" />
  </xs:restriction>
</xs:simpleType>
```

Chapter 6. GHSTS Package Definition

This chapter describes the GHSTS package, which is the physical representation of a GHSTS submission.

6.1. Overview of Folder structure

The following table shows the folder structure of a GHSTS submission. As an example the displayed submission is the submission with submission number 03 of a series of submissions from 01 to 04.

Parent hierarchy	Regulatory action level	Submission level	Name	Description (Bold = name of folder, as used in description of submission package)
				Folder hierarchy above dossier level (unknown and out of scope for this specification).
			SYNCH127 (example)	Regulatory action folder Root folder for the regulatory action / dossier. Only submissions of this dossier allowed. A GHSTS viewer does not need to access content outside of this regulatory action folder.
			01	Submission folder Root folder for submission 01
			02	Submission folder Root folder for submission 02
			03	Submission folder Root folder for the submission 03.
			main.ghsts	GHSTS XML backbone file of this submission.
			main.html	static human-readable representation, as generated by the XSLT (see chapter 2.4.1.9)
			content	Content folder Folder for all computer files that are referenced by a FILE.
			utils	Utils folder Folder for additional files relevant for display or validation.
			ghsts_02-00-00.xsd (example)	XSD (XML schema definition) file of the GHSTS XML backbone in the submission folder. Note: The filename is an example and may change when versioned.
			CommonComponents.xsd	XSD for the CommonComponents XSD (see chapter 5.1.8) which is referenced by the GHSTS XSD
			toc	ToC folder If the GHSTS backbone file references a standard ToC XML file then this file and its Toc XSD file is placed in this folder.
			resources	Resources folder Folder for additional resources needed for display and validation of individual content files, e.g. style sheets, XSD or other assets to display and eventually validate XML summary files in a GHSTS viewer. The folder may contain subfolders, e.g. to group assets by OHT template.


	04	Submission folder Root folder for submission 04
---	----	---

Table 103 – overview of folder structure

There is no separate folder for auxiliary information, for example side letters. All content information supplied by the registrant in the submission package must be put into the content folder and referenced from a FILE instance in the GHSTS XML backbone file. This means that only content can be submitted in a submission package that is referenced from a ToC node.

Please note that that submission package does not contain a representation of the picklist model, that was used to compile the submission. The information in the picklist item type is self-contained and sufficient to display the decode values for one language. In order to display other language values in a GHSTS viewer or to validate the used picklist values against the picklist model in a GHSTS validation tool, those components need to use a representation of the picklist model themselves. The build information (see chapter 4.9.5) in the XML backbone may contain additional information about what picklist model was used to compile the submission package.

6.1.1. Usage of Content folder

The actual content of the submission, which is the referenced computer files of the FILE instances in the GHSTS XML backbone, is located in the content folder. For file formats see chapter 3.2.15.

Each referenced computer file is contained once in the submission package, even if the document to which it belongs is assigned to multiple ToC nodes / data points at the same time. Each referenced computer file is stored in or below the content folder. It is not allowed to use file system functions like links in Windows or Linux to mimic these multiple assigned documents as multiple computer files on the file system.

The content folder may contain subfolders however this is only recommended when the number of computer files in the content root folder becomes too large to be efficiently displayed in a file system browser. It is not recommended to create a copy of the hierarchical ToC structure in the form of folders – as outlined above this would come to its limits for multiple assigned documents anyway. The source for the ToC node / data point information is the GHSTS XML backbone file, not the folder structure below the content folder.

6.2. Storage of multiple submissions of the same dossier

As discussed in chapter 3.2.29 the GHSTS XML backbone file of a submission can contain two types of external references that point to predecessor submissions:

- References with PID.
- For incremental submissions: file references to computer files in predecessor submissions (see chapter 3.2.22)
The reference from the GHSTS XML backbone file to the computer files in the main and attachments folders is a relative path that always first goes up one level to the submission folder and then descends to the correct submission folder (see chapter 6.2). By this way the structure of the file path is always identical for full and incremental submissions where the content is located in a previous submission.

If the receiver work with a file system-based dossier management system then the submissions must be stored below the same regulatory action folder as outlined in **Table 103** in order that these references can be resolved programmatically by a GHSTS viewer or another retrieval component working on a file system level.

The registrant has to – at least temporarily – store the submission in the same way to be able to check the consistency of the submission with the GHSTS validator that will need to resolve and verify the external references.

The total length of a relative path is restricted to 230 characters. Due to a restriction on some Windows file systems where the total absolute path must not exceed 255 characters it is advisable to create the regulatory action folder not too “deep” below the drive letter on these systems.

6.3. Folder naming conventions and constraints

The following table lists the folder naming conventions and constraints for the folders listed in the figure above.

Folder type	Folder naming convention and constraints	Example
Regulatory action folder	Folder is outside of scope of GHSTS submission package and will only exist if the regulatory authority uses a file system-based dossier management system. Recommendation for folder name: Value of DOSSIER_COMP_ID of the DOSSIER element – this value should be identical in all submissions of the regulatory action.	SYNCH127
Submission folder	Folder name is submission number which is padded to two digits. Folder must only include the files and folders as shown in table “Table 103 – overview of folder structure”.	01 02 03
Content folder	Mandatory folder name is “content” The main folder must contain all computer files of FILE instances May contain subfolders	content
Utils folder	Mandatory folder name is “utils” May contain subfolders	utils
ToC folder	Mandatory folder name is “toc” May not contain subfolders	toc
Resources folder	Mandatory folder name is “resources” May contain subfolders	resources

Table 104 – folder naming conventions

6.4. File naming conventions

The following table lists the file naming conventions or recommendations that exist for the computer files that are part of a GHSTS submission package.

Purpose of computer file	Parent folder	Naming convention	Example
Static human-readable representation of the contents of the GHSTS XML file	Submission folder	Filename 'main.html' is mandatory.	main.html
GHSTS XML backbone file of this submission	Submission folder	Filename “main.ghsts” is mandatory. The suffix “ghsts” can be used to be registered in the operating system in order to automatically launch a GHSTS viewer.	main.ghsts

GHSTS XSD that describes the structure of the submitted GHSTS XML backbone file	Utils folder	See chapter 4.2, do not alter supplied filename	ghsts_01-00-00.xsd
XSD file of the CommonComponents XSD referenced by the GHSTS XSD	Utils folder	See chapter 5.1.8, do not alter supplied filename	CommonComponents.xsd
XSD file of the Standard ToC file	Toc folder	See chapter 5.1.2, do not alter supplied filename	toc_01-00-00.xsd
XML file of a Standard ToC	Toc folder	See chapter 5.1.2, do not alter supplied filename	toc-oecd_micro_01-00-00.xml
Computer file in content folder (or subfolder)	Content folder	<p>There is no binding naming convention for files in the content folder, as long as the file names are unique per folder. Please note that a filename must not contain all possible characters – see chapter 5.2.7 – and illegal characters must be replaced by a legal character, e.g. “_”.</p> <p>Suggestion for naming convention (non-binding):</p> <p>As every computer file belongs to exactly one document a short human readable document identifier like the company study number in the element DOCUMENT_NUMBER could be used. If multiple languages are submitted simultaneously the language code of the study is added. Alternatively the company can use another internal identifier that identifies the document uniquely across different languages.</p> <p>This file name part is then followed by a suffix indicating the file type of the document (e.g. “main”, “source” and “supp”) and for file types (e.g. “Main”, “Source” and “Supplemental”). If there are multiple computer files of the same file type a number could be added.</p>	<p>C6313_0016-EN-main.pdf</p> <p>C6313_0016-EN-source.doc</p> <p>C6313_0016-EN-suppl.bmp</p> <p>C6313_0016-EN-suppl2.bmp</p>

Table 105 – file naming conventions

6.5. Platform and media support

The definition of a GHSTS submission package is not bound to a specific computer platform, file system type or media. A GHSTS submission package can be created and stored on any standard platform, arbitrary media or transport protocols can be used to transport it from the registrant to the receiver(s).

GHSTS does not contain mechanisms to detect alterations of the submission package itself that may arise during its transport. On the content level the MD5 checksum of the FILE content can help to verify the integrity of the files in the content folder.

Chapter 7. Using GHSTS

This chapter describes in a brief manner, how GHSTS can be adopted and used by registrants

7.1. OECD GHSTS website

Information about the current status of GHSTS and its components are available on the OECD website at: <http://www.oecd.org/chemicalsafety/submission-transport-standard/>.

On the OECD website the user will find

- the released versions of the GHSTS core components for download.
- information about additional GHSTS components.
- further information, like manuals.

7.2. Creation and publication of a GHSTS submission

To get ready to create a first GHSTS submission, the registrant has to do the following preparation steps:

- Enable in-house dossier management system to support equivalents of the main concepts of GHSTS:
- Document Families, Documents, Files.
- Notion of content changes for documents as described in chapter 3.2.18.
- Notion of ToC node assignment status
- Notion of metadata changes
- Generation and assignment of local PID.
- Submissions, being able to supply incremental submission as comparison to previous submissions.
- Import available GHSTS picklists and their values into in-house system to use for metadata assignment, allow repeated update of values.
- Import available predefined ToC templates into in-house system to use for dossier creation, allow repeated update of structures. Please note that the special characters < (<), & (&), > (>), " ("), and ' (') are encoded in the XSD file and need to be decoded before usage in the in-house system.
- Create GHSTS publication component as new output format for dossier management system.

Eventually the use of a converter from an existing standard to GHSTS can serve as intermediate and temporary step. However, it is very likely that existing standards do not incorporate all mandatory information in a GHSTS submission package.

7.3. Technical Validation

Before a GHSTS submission is submitted to the recipient, the compliance with the GHSTS standard has to be verified using a GHSTS validator. To do so the sender can apply a technical validation ruleset (see chapter 2.4.1.10). This validation ensures that a submission package can be processed correctly by the recipient. It does not contain any validation rules that are specific to a regulated domain or regulation but checks conformity with the underlying business logic, e.g. any dependencies between values.

This technical validation may be complemented by further domain, regulation or authority-specific validations. Their definition and application is out of scope for GHSTS.

Optionally, the sender can supply information about the latest validation below the XML element GHSTS/SENDER_ENVIRONMENT/VALIDATOR.

7.4. Transport of a GHSTS submission

A GHSTS Package can be transferred to the authorities on any kind of electronic media or with any kind of electronic protocol depending on the agreement with the regulatory authority. The regulatory authority is responsible to copy the submission into the file system or into their in-house dossier management system.

7.5. Consumption of a GHSTS submission

There are three scenarios for the consumption of a GHSTS submission by the receiver:

- The receiver uses a GHSTS viewer to open a submission.
- The receiver uses an in-house dossier management system that possesses comparable concepts as within GHSTS.
- The receiver can use the GHSTS submission package in a “standalone” manner, using the static HTML representation to view basic the information in the backbone file and the file content. This scenario is only advised when the first two options are not available.

For the second scenario the receiver will require a specific GHSTS ingestion component. This component will parse the GHSTS XML backbone and compare the supplied information with the data that is already stored internally, for example for PID of files and documents. The ingestion may then be only partially dependent on the supplied information for content and metadata changes.

Chapter 8. Appendix

8.1. Glossary of Terms

The definitions used in this document are for the purpose of the GHSTS only.

Terms	Definitions
ACVM	See “MPIMAF-ACVM”.
Annex Point	A colloquial short cut for specific required information defined in a regulation / legislation. Annex Point could be used as a synonym for a (leaf) TOC-Node in a TOC.
Applicant	Company, “leading company” of a shared dossier of multiple companies, or individual that is asking for approval for use and/or marketing a certain component or product to a regulatory authority.
Application	See “Dossier”.
Authority	A global, regional or national governance body that establishes, monitors, reforms, and enforces regulations in a specific regulated domain (e.g. pesticides), i.e. Regulatory Agency, Regulatory Authority.
BVL	German Federal Office of Consumer Protection and Food Safety Responsible for the authorisation of plant protection products, veterinary drugs and genetically modified organisms.
CADDY	Computer Aided Dossier and Data supply CADDY is an electronic format for the exchange, archiving and evaluation of dossiers, developed jointly by industry and regulatory authorities. The current standard CADDY-xml has similar characteristics as GHSTS and can be considered its development predecessor. Citation: ECPA eSEG (Electronic Submission Expert Group), see http://esubmission.ecpa.eu/caddy .
CAS Registry Number	A unique numeric identifier generated by Chemical Abstracts Service designating a substance. The number is a numeric identifier that can contain up to 10 digits, divided by hyphens into three parts (e.g. XXXXX-XX-X). CAS Registry Number is a Registered Trademark of the American Chemical Society.

Terms	Definitions
CBI	Confidential Business Information - Since the distinction of what is CBI and what is not tends to vary from country to country, it was determined that the harmonised schema would need to allow for multiple CBI fields which would indicate the CBI status of a document across multiple regulatory authorities. Further distinctions between the documents can be made using specific Document IDs. Confidential Business Information can generally be described as information (trade secrets or information of commercial value) to concern the following: - manufacturing or quality control processes relating to an active substance or pest control product; or- methods for determining the composition of an active substance or pest control product; or- the identity and concentration of the formulants and contaminants in a pest control product, other than those considered to be of health; or - environmental concern that are identified on a list to be made available to the public.
CFR	US Code of Federal Regulations; legal text containing the United States Environmental Protection Agency pesticide regulations.
Dossier numbering systems (see also: ToC: - Table of Content):	System of codes used to identify the types of documents included within the dossier. These codes can identify letters, forms, data, reviews, and other document types. Examples are shown below. Other numbering systems exist in different regions and authorities: Organisation For Economic Co-operation and Development (OECD) IIA (active substance) /IIIA (pest control product) numbering system PMRA (DACO): Canada - Pest Management Regulatory Agency (PMRA) uses the Data Codes numbering system (DACO) USEPA (OPPTS): United States - Environmental Protection Agency (EPA) uses the OPP & OPPTS system
Decision	An authoritative determination made after consideration of facts from a submission, document, or report.
DER	Data Evaluation Record; a review of data from a specific document or data point used by PMRA and EPA. The DERs are summarized to provide the final review of the data.
Document	The written content of study reports, labels, forms, cover letters, etc. which are created by the registrant. A Document is a primary submission component and may be subject to amendments. A document contains of at least one file as “Main” representation and may have additional complementary files as “Attachments”. The “Attachments” share the same lifecycle as the “Main” representation”.
Document family	A Document Family is a logical binding of Documents containing the same information (e.g. study report) in different versions (e.g. study report amendments and variations (e.g. translations, regulatory-specific study report layout variations like stamping and signature page).
DOI	DOI Digital Object Identifier DOIs uniquely identify digital objects and are widely used to identify academic, professional, and government information. (https://www.doi.org/)
Dossier	The set of documents prepared by the applicant submitted for a regulatory purpose. It is maintained continuously throughout the lifecycle of a regulatory action. A product may be composed of a number of dossiers, such as the initial dossier to register the product, as well as subsequent dossiers to amend the product. A Dossier may also have numerous ‘submissions’, which are used to submit documents to the Regulatory Authorities either as part of a new dossier or an existing dossier. One Dossier = one Regulatory Action = including maximum one product

Terms	Definitions
Dossier, lifecycle management	The coordination of all versions of submissions, documents, and files throughout the regulatory submission process from submission to approval of a regulatory action.
EC/European Commission	The European Commission is the executive body of the European Union (EU).
ECHA	European Chemicals Agency manages the technical, scientific and administrative aspects of REACH and is also the owner of the IUCLID database software.
e-Dossier	The e-Dossier is the electronic representation of the dossier.
e-Index Builder	Produced by the Canadian Pest Management Regulatory Agency (PMRA). The e-Index Builder is a stand-alone Java-based software application which facilitates the creation of an electronic index (in XML format) that fully describes each document submitted in support of applications to register or amend a pest control product registration, re-evaluation, special review, or in response to requests for information. Citation: Health Canada (n.d.), Health Canada website, www.hc-sc.gc.ca/cps-spc/pest/registrant-titulaire/eprs-serp/_eib-gie/index-eng.php (accessed 17 September 2013).
EPA (New Zealand)	New Zealand's Environmental Protection Authority (EPA) administers applications for major infrastructure projects of national significance, and regulates new organisms (plants, animals, GM organisms) and hazardous substances and chemicals. The EPA also administers the Emissions Trading Scheme and New Zealand Emission Unit Register, and manages the environmental impact of activities in New Zealand's Exclusive Economic Zone (EEZ), including prospecting for petroleum and minerals, seismic surveying and scientific research. Reference: Environmental Protection Authority of New Zealand (n.d.), http://www.epa.govt.nz , (accessed 17 September 2013).
EPA (US)	Environmental Protection Agency is an agency of the United States Federal Government charged to protect human health and environment.
e-Prism	US EPA multi-faceted system for consolidation and management of EPA pesticide information. This system serves as a gateway to electronic submission of pesticide-related information and includes a portal for XML based electronic submissions.
e-PRS	Canada's Electronic Pesticide Regulatory System for consolidation and management of PMRA information. This system serves as a gateway to electronic submission of pesticide-related information and includes a portal for XML based electronic submissions.
e-Submission	The electronic representation of a submission to an authority according to agreed standards including lifecycle management.
European Union (EU)	The economic and political union of 28 member states located in Europe.
File	Files – the physical manifestation of a document's written content (or part of the content) in the form of a currently accepted formats.
GHSTS	OECD Global Harmonised Submission Transport Standard provided to Companies and Regulatory Authorities as the standard for the submission of e-Dossiers to regulatory authorities for the review of pest control products.

Terms	Definitions
GIFAP	International Group of National Associations of Manufacturers of Agrochemical Products (from French: Groupement International des Associations Nationales de Fabricants de Produits Agrochimiques, GIFAP). In 1996, GIFAP was renamed to Global Crop Protection Federation (GCPF). In 2001, GCPF was renamed to CropLife International.
Global Review	A formal process with negotiated timelines for the review of a dossier, where the workload is split among participating regulatory authorities globally, and the reviews of data are exchanged and peer-reviewed.
IANA	Internet Assigned Numbers Authority coordinates globally Internet protocol resources (https://www.iana.org)
IUCLID	International Uniform Chemical Information Database A software application to capture, store, maintain and exchange data on intrinsic and hazard properties of chemical substances. It is the key tool for chemical industry to fulfil data submission obligations inside the REACH legislation. The software is maintained by the European Chemicals Agency. Reference: European Chemicals Agency (n.d.), IUCLID website , https://iuclid6.echa.europa.eu/ (accessed 31 January 2019).
Joint Review	A registration dossier package to be reviewed, either independently or dependently, by two or more regulatory authorities from two or more different global regions. The evaluation of a pesticide dossier is shared by two or more countries. The participating regulatory authorities review the work of the primary reviewers for each particular science discipline, and the end product (ideally a complete monograph or key components of the monograph) is used by all participating countries (and others) as the basis for regulatory decisions.
MPIMAF-ACVM	Ministry for Primary Industries Forestry - Agricultural Compounds and Veterinary Medicines Responsible for registration of agricultural compounds (including pesticides) and setting of Maximum Residue Limits, along with the development, implementation, evaluation and reviews of standards relating to agricultural compounds.
Metadata	“Data about data”. The information fields or attributes describing with keywords the content of a document or dossier.
Monograph	An evaluation report issued by the authorities.
MRID	Master Record Identification Number A unique cataloguing number assigned to an individual pesticide study at the time of its submission to the US EPA. Citation: United States Environmental Protection Agency (n.d.), Pesticides: Glossary website, www.epa.gov/pesticides/glossary/m-q.html (accessed 17 September 2013).
Multinational Submission	A registration application and associated content which is submitted for review to two or more countries.
Numbering system	See "Dossier numbering system".

Terms	Definitions
OECD	The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation which includes 36 member countries committed to democracy and the market economy. Its principal aim is to promote policies for sustainable economic growth and employment, a rising standard of living and trade liberalisation. By “sustainable economic growth” the OECD means growth that balances economic, social and environmental considerations.
OECD Expert Group on the Electronic Exchange of Pesticide Data	A group of experts from OECD governments, industry and other stakeholders, with expertise in approaches and/or issues associated with IT systems and pesticide registration, established by the OECD to identify and address issues to facilitate the efficient electronic exchange of pesticide information in the pesticide regulatory process. The work of this group is overseen by the OECD Registration Steering Group.
OECD Expert Group on Harmonising Templates	Group of experts from OECD member countries, industry and NGOs, in charge of the development and review of OECD Harmonised Templates for Reporting Test Summaries.
OECD Harmonised Templates	An OECD harmonised template is a standard data format for reporting studies done on chemicals to determine their properties or effects on human health and the environment. They are aimed at developers of database systems, as they prescribe the formats by which information can be entered into and maintained in database. By using these templates, governments and industry will easily be able to electronically exchange test study summary information. The templates can be used for reporting summary results for testing on any type of a chemical (e.g., pesticides, biocides, industrial chemicals).
OECD XML study templates	The XML format of an OECD harmonised template.
Pesticide Package	The collective contents of the pesticide registration application, documents/studies, label(s), and any other materials needed to initiate a regulatory review process.
Pesticide Regulatory Process	The process handling the registration in accordance with the government's pesticide regulation laws.
Pesticide Review Process	The process of assessing the validity and conclusions within a registrant's application package, including the examination of all studies and summaries, as well as the proposed product label.
PID	Persistent identifier
PMRA	Health Canada's Pest Management Regulatory Agency (PMRA) registers all pesticides that are sold, used, manufactured or imported into Canada under the federal <i>Pest Control Products Act 2002 c.28 (PCP Act)</i>
PR Notice 86-5	The EPA-issued notification describing the standard format for data submitted under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) and certain provisions of the Federal Food, Drug, and Cosmetic Act (FFDCA). This notice has now been superseded by PR Notice 2011-3.
PRIA	US Pesticide Registration Improvement Act which defines the review time and fee for service for pesticide reviews.

Terms	Definitions
Primary Reviewer	AKA Lead Reviewer. The evaluator from the regulatory authority assigned the lead for each science discipline who does the initial data review and drafts the primary review (study evaluation).
Project	In the e-dossier system, project is designated by the registrant and has a defined start and end.
PRZ	The file extension assigned to the output file from the use of the PMRA e-Index Builder.
RA	Regulatory Authority
Rapporteur	In the context of dossiers, the EU member state who is the primary reviewer of a submitted registration application and who shares their findings with the remaining EU member states.
REACH	Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. A European Union Regulation which addresses the production and use of chemical substances. Reference: European Union, 1998-2013, EUR-Lex website, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32006R1907:en:NOT (accessed 17 September 2013).
Reference List	The list of documents and studies submitted by a registrant as part of a registration submission package.
Registrant	Company, company representative, or individual who holds the registration for a product, e.g. in the pesticides domain an active substance or pest control product). Also called “Approval Holder”.
Registration	The process to apply, assess and register a specific component (e.g. pesticide).
Regulation	Legal act
Regulatory action	See “Dossier”.
Regulatory Agency/ Regulatory Authority (RA)	See “Authority”.
Regulatory Process	The process handling the registration in accordance to the regulation.
Review Process	The process of assessing the validity and conclusions within a registrant's application package, including the examination of all studies and summaries, as well as the proposed product label.
Reviewer	A regulatory authority employee or contractor involved in a review process, having the responsibility to ascertain the validity and conclusions of a registrant's application package.
Specification	A set of standards for constructing an electronic dossier, which if compliant, enables the submission of said compliant package to an authority for consumption into a database or other organized system.

Terms	Definitions
Study Report	Is a document which reports the results of data from laboratory or field tests. In most cases it is conducted under good laboratory practice (GLP) conditions.
Test Summary	A Test or Study Summary provides key information about a study in a condensed form. Study summaries may be presented in text files or in a database representation (e.g. OECD Harmonised Templates).
Style sheet	Style sheets are the way that standards-compliant Web designers define the layout, look-and-feel, and design of their pages. They are called Cascading Style Sheets or CSS.
Submission	A compilation of documents prepared by the applicant for a specific regulatory purpose in a structured form according to the given regulatory requirements. Multiple submissions can be submitted for each dossier.
Submission, Version of -	A dossier, submitted with the goal of achieving a positive regulatory decision, can consist of multiple deliveries of documents or data toward that goal. Each delivery, termed a submission, is assigned a version number for tracking purposes.
Submitting Entity	The name of the registrant company that submits the registration application and accompanying data.
Table of Contents	Represents a hierarchy for documents within the dossier. For different legislations different Table of Contents (TOCs) exist.
Template	A pre-developed layout in electronic or paper media .An electronic file with a predesigned, customized format and structure, as for a fax, letter, or expense report, ready to be filled in.
TOC	See "Table of Contents".
TOC Node	One node in the tree structure of a TOC
Transport Mechanism	Systems designed for handling large numbers of documents and for exchanging such documents electronically in a well-defined and standardized framework.
Transport Mechanism Subgroup	Subgroup of the OECD Expert Group on Electronic Exchange of Pesticides Data (EGEPPD).
UUID	Universally Unique Identifier: This identifier enables distributed systems to uniquely identify entities, such as legal entities, products, dossiers, components, documents, files. For more details see RFC 4122: A Universally Unique Identifier (UUID) URN Namespace or Wikipedia.
Validation	The establishment of documented evidence that a system does what it purports to do, and the processes by which the reliability and relevance of a particular approach, method, process or assessment is established for a defined purpose.
Working Group on Pesticides	Working group that directs and oversees the work of the OECD Pesticide Programme, a forum for national pesticide regulators from OECD member and selected non-member countries to discuss and find solutions to issues of common interest.
Work sharing	A formal process among regulatory agencies to share the work to review a dossier.

8.2. Standards References

The following table contains a list of references to official standards that are mentioned in this specification:

References	Description
ISO 3166-1-alpha-2	see http://www.iso.org/iso/iso-3166-1_decoding_table.html
XML	see http://www.w3.org/XML/
XML schema	see http://www.w3.org/XML/Schema.html#dev
RFC 1321	MD5 Message Digest Algorithm, see http://www.ietf.org/rfc/rfc1321.txt
RFC 2141	URN syntax, see http://tools.ietf.org/html/rfc2141
UUID	UUID syntax see http://www.itu.int/rec/T-REC-X.667-200808-1/en
ISO-19005-1:2004	Document management – Electronic document file format for long-term preservation - Part 1: Use of PDF 1.4 (PDF/A-1), ISO, 2004

Table 106 – standards references

8.3. Copyright and Trademark Information

- The copyright for this document is with the OECD.
- The figures in the text are created using the Open Source software yED (http://www.yworks.com/en/products_yed_about.html).
- The graphical representations of the XSD schemata are created using Altova XML Spy 2013.
- The icon for the Adobe PDF document is a trademark of Adobe Systems Incorporated.
- The icons for Microsoft Word and Microsoft Excel documents are trademarks of Microsoft Cooperation.
- The icons used in the figures come from various sources:
 - Hewlett Packard Icon Database from the HP Experience Center. Copyright Hewlett Packard, released under license.
 - Picol Library by Melih Bilgil, see <http://www.picol.org>, licensed under the Creative Commons-License BY-SA (<http://creativecommons.org/licenses/by-sa/3.0/>).
 - Lab icon from <http://www.wpzoom.com>, licensed under the Creative Commons Attribution-Share Alike 3.0 unported License (<http://creativecommons.org/licenses/by-sa/3.0/>).
 - Format icons from the icon set of the Regional Computing Centre of Erlangen, based on the Tango desktop project, licensed under the Creative Commons Attribution Share-Alike 3.0 license (<http://creativecommons.org/licenses/by-sa/3.0/>).
 - Folder and document icon from Momentum Designlabs, see <http://momentumdesignlab.com> licensed under the Creative Commons Attribution-Share Alike 3.0 license (<http://creativecommons.org/licenses/by-sa/3.0/us/>).

8.4. Version history

The following Change History log contains a record of changes made to this document. The last column also contains information whether a new version of this GHSTS specification also requires a change of other GHSTS components (e.g. the GHSTS schema definition).

Version #	Date	Author	Section / Nature of change / Affected GHSTS components
01.00.00	May 2014	OECD	Initial publication of this document.
02.00.00	June 2019	OECD	Major update of all GHSTS components to support the use of the GHSTS for electronic packages in other regulatory domains in addition to the pesticides domain and to include improvements in the lifecycle management of documents.

Table 107 – version history

---- End of document ----