COLLEGE OF COMPUTING, INFORMATICS AND MATHEMATICS
UNIVERSITY TEKNOLOGI MARA

CDIM263 - BACHELOR OF INFORMATION SCIENCE (HONOURS) INFORMATION CONTENT MANAGEMENT

**INTRODUCTION TO METADATA
(IMC 452)**

**GROUP:**
CDIMST2

**INDIVIDUAL ASSIGNMENT:**
POLICY

**PREPARED FOR:**
TS. DR. AP-AZLI BIN BUNAWAN

**PREPARED BY:**

| NAME | MATRIX NUMBER |
|---|---|
| MUHAMMAD IQBAL BIN BAHARI | 2024699752 |

**SUBMITTED ON:**
13 JUNE 2025

**ACKNOWLEDGEMENT**

I would like to express my sincere gratitude to all those who have contributed to the completion of this assignment. My deepest appreciation goes to my lecturer, Ts. Dr. Ap-Azli Bin Bunawan, for the continuous guidance, valuable insights, and unwavering support throughout the development of this study. I am also thankful to University Teknologi Mara for providing the necessary resources and a conducive academic environment for my assignment. Special thanks to my colleagues and peers for their constructive feedback, encouragement, and assistance during various stages of this work. Finally, I extend my heartfelt appreciation to my family and friends for their patience, motivation, and moral support throughout this academic journey.

**TABLE OF CONTENT**

## TYPES OF METADATA:

### 1. DESCRIPTIVE METADATA

Descriptive metadata represents the most widely recognized category of metadata, focusing on identifying and describing the intellectual content of digital resources. Through descriptive metadata users can discover resources while also understanding their content characteristics to evaluate their value (Adamu Ahmed et al., 2024). Descriptive metadata includes title information and author names and subject classifications and abstracts and keywords and other identifying details which help users assess resource value and relevance.

The main purpose of descriptive metadata exists to facilitate resource discovery and retrieval operations. Descriptive metadata functions as the connection between users' information requirements and the extensive digital content collections in digital libraries and repositories (Glowacka-Musial, 2022). The process of subject analysis within descriptive metadata creation requires intellectual evaluation to select appropriate subject headings and classification terms which precisely represent document or digital object content. The traditional manual process of subject analysis has received improvements through automated methods which include topic modeling and natural language processing (Glowacka-Musial, 2022).

Good descriptive metadata has great influence on users and efficiency of a system. Studies have revealed that more than 84 percent of academic planners and scholars are unhappy with the quality of metadata available today with a sizeable number describing it as poor and unusable to their satisfaction (Pro-Metrics, 2024). The source of this dissatisfaction is the high prevalence of errors during metadata creation, the irregular use of standards, and the inability to cover all the aspects of the resources with these standards. The use of standard vocabularies and controlled terms, including ORCID identifiers and DOI systems, has become an urgent way to solve the problem of the low quality of descriptive metadata (Pro-Metrics, 2024).

Artificial intelligence and machine learning methods are becoming more commonly used in modern descriptive metadata creation methods. Large Language Models (LLMs) have been promising automation of metadata enrichment workflows, and ChatGPT and GoogleGemini have proven to be better at topic classification tasks than their conventional counterparts (Martorana et al., 2024). Such technologies hold a promise of decreasing the manual workload in creating metadata with high accuracy and consistency rates.

Development of descriptive metadata standards are still evolving to meet the changing technology scenes and user requirements. It includes the semantic web technologies integration, such as Simple Knowledge Organization Systems (SKOS) and Resource Description Frameworks (RDF), which offer better interoperability and machine-readability (Hancock, 2020). Such advances allow more advanced search and discovery applications and provide the increasing requirement of linked data and semantic integration between various information systems.

## 2. STRUCTURAL METADATA

Structural metadata describes the internal structure and interrelationships of digital objects and is the architectural plan that allows digital complex resources to be properly rendered, navigated and preserved (Mosha & Ngulube, 2023). This type of metadata is especially important to compound digital objects, which are made up of many parts, such as books and chapters, websites and linked pages, multimedia presentations and synchronized components.

The fundamental purpose of structural metadata is to maintain the integrity and usability of digital objects across different systems and over time. It specifies the relationships between individual components, and their hierarchical structures, and the technical details needed to ensure correct display or playback (Xie et al., 2019). In the case of scientific data preservation, structural metadata is crucial in making certain that complicated datasets will continue to be available and understandable to future research activities. The metadata records essential data on data formats, file dependencies and transformation procedures which could be required in long-term preservation plans.

Metadata standards have been developing towards the greater complexity of digital objects and the requirement of interoperability among different systems. An example of this development is the Metadata Encoding and Transmission Standard (METS), which defines a way of encoding descriptive and structural metadata in the same schema (Formenton & Souza Gracioso, 2022). METS allows institutions to build self-contained metadata packages which may be used as Open Archival Information System (OAIS) information packages so that the digital objects can be accessible and usable in the long term.

Structural metadata may only be implemented with great attention to both technical specifications and preservation requirements. An important part of digital preservation, format migration depends on structural metadata to support the idea that the transformations between one file format and another will preserve the key properties and associations of the initial objects (Xie et al., 2019). It is done via a complicated set of mapping rules that describe how structural elements are to be carried over or transformed in format conversions, such that the intellectual content is not lost due to changes in the underlying technology.

Recent structural metadata systems are starting to add machine-actionable representations which can be processed and transformed automatically. The emerging complexity of structural metadata applications is evidenced by the development of specialized languages and schema, such as in statistical data processing (Alter et al., 2020). These structures allow intricate operations such as merging, reshaping, and transforming of data whilst keeping extensive records of any structural changes and relationships.

## 3. ADMINISTRATIVE METADATA

Administrative metadata consists of management information required in the lifecycle management of digital resources, such as technical specifications, preservation actions, rights management and access control information (Mosha & Ngulube, 2023). It is a type of metadata that is the operational backbone of the digital asset management systems and allows organizations to manage, track and preserve their digital collections in an effective manner.

Administrative metadata is wider than mere cataloging; it is an extensive documentation of all administration procedures and technical features that relate to the digital objects. This includes creation dates, file sizes, checksums for integrity verification, migration histories, and audit trails of all preservation actions (Steeman & Schuurman, 2018). The PREservation Metadata: Implementation Strategies (PREMIS) standard is a framework in which administrative metadata can be organized and implemented, and it supplies recommendations on how the many kinds of preservation-relevant information can be grouped and defined.

An important use of administrative metadata is authentication and integrity checking. Administrative metadata records digital signatures, checksums and other cryptographic measures that can be used to prove the authenticity and integrity of digital objects over time (Xie et al., 2019). The technical safeguards are the more significant the digital objects are migrated between systems, transformed in terms of formats, or used by multiple persons on various platforms.

Administrative metadata implementation needs institutional policies and technical infrastructure to be considered closely. Organizations must develop customized metadata dictionaries that align with their specific preservation goals and operational requirements (Steeman & Schuurman, 2018). The experience of the Netherlands Institute for Sound and Vision in creating working standards for preservation metadata illustrates how institutions can tailor a set of general guidelines to their own requirements and still achieve interoperability with the wider preservation communities.

Automated capture and management are being added to modern administrative metadata frameworks. High-performance computing environments like EMPRESS show how administrative metadata can be used to speed scientific discovery by selective access to data by its descriptive properties (Lawson et al., 2022). Such systems provide such features as scaleable metadata consistency mechanisms, relaxed system configurations, and fault-tolerance mechanisms tailored to production-worthy digital preservation systems.

## 4. PRESERVATION METADATA

Preservation metadata is a particular kind of administrative metadata devoted solely to the long-term accessibility and usability of digital objects (Mosha & Ngulube, 2023). It is a type of metadata that records all the actions, decisions and technical requirements needed to preserve digital resources over time, usually decades or centuries.

The overarching problem that preservation metadata is designed to solve is that it is not known what future technology will require, nor what uses will be demanded, and therefore the maximum possible documentation must be retained (Marjolein Steeman & Schuurman, 2018). The metadata used in preservation should not merely reflect the current technical standards but should also include some contextual data regarding how the object was created, what software it relies on, and what kind of environmental conditions it might be sensitive to (Xie et al., 2019). This future-oriented model demands institutions to consider the possible preservation issues in advance and write down the information that can be of primary importance in relation to future migration or emulation plans.

The implementation of preservation metadata standards requires careful balance between comprehensiveness and practicality. The conceptual model upon which preservation metadata requirements can be considered is the Open Archival Information System (OAIS) model, which is accompanied by the practical implementation guidelines in the form of the PREMIS (Formenton & Souza Gracioso, 2022). The standards promote the value of recording provenance data, technical descriptions, and preservation activities in a form that will be meaningful andenable action over time scales of years.

Format migration is one of the most important phenomena of preservation metadata, because it deals with the migration of digital objects between one format and another, preserving their key properties (Xie et al., 2019). It is insufficient that preservation metadata records the technical details of the source and target formats: the rules by which similarities between them are mapped, and the transformations undertaken during migration, also need to be recorded. This documentation will be vital in justifying success of migration processes and future preservation actions planning.

The effectiveness of preservation metadata frameworks depends heavily on institutional commitment and technical expertise. Studies found that the quality of metadata degrades over time and that the preservation information needed is reduced significantly each year (Crandall et al., 2022). The significance of this metadata decay phenomenon is that it emphasizes the value of active preservation metadata management and necessity of constant monitoring and refreshing of preservation documentation.

## 5. RIGHTS METADATA

Rights metadata Rights metadata consists of legal and intellectual property data relating to digital resources, such as the copyright status, usage rights, license information, and access controls (Matusiak et al., 2021). This form of metadata acts as the legal application within which the digital resources could be utilised, shared, and re-distributed and is, therefore, crucial to adhering to intellectual property legislation, as well as institutional rules and regulations.

Rights metadata can be complex due to wide variety of legal frameworks and licensing agreements under which digital content is published. The problem of copyright evaluation of archive materials has its specifics as there is often little information on authorship, dates of creation, or ownership history (Matusiak et al., 2021). Community archives, often operated by volunteers, face additional difficulties in accurately assessing and documenting rights information, leading to potential inaccuracies in rights statements and usage permissions.

Controlled vocabularies of rights have become recognized as important to enhancing the predictability and machine-readability of rights data. RightsStatements.org is a controlled vocabulary of standard, machine-operable statements about the copyright status of digital resources, jointly created by the Digital Public Library of America (DPLA) and Europeana (Ballinger et al., 2017). These standardized statements allow cultural heritage institutions to give clear and consistent information about usage rights and they also allow automated processing and aggregation of rights information across multiple repositories.

Realization of the rights metadata needs to give precise attention to the legal experience as well as the establishment policies. Lots of libraries and cultural heritage organizations do not have any specialized copyright expertise, and it is very difficult to evaluative and record rights information precisely (Ballinger et al., 2017). Formulation of practical guidance materials and education resources to assist institutions to make suitable rights decisions and prevent possible legal entrapments has become imperative.

In the context of rights management, modern systems are augmented with blockchain technology and smart contracts to give an automated process of enforcing rights and distributing royalty (García et al., 2023). These systems allow to create an immutable record of rights ownership and usage rights permissions and facilitate automated performance of licensing agreements. Metadata about rights linked to blockchain systems promises to provide a solution in situations where the rights management is challenging in digital media and entertainment sectors (Klaris, 2015).

## 6. STATISTICAL METADATA

Statistical metadata is the structured information describing statistical properties of data, the processing procedures and quality aspects of statistical data, helping statistical information users to comprehend the information and use it adequately (Signore et al., 2015). This type of metadata plays a very important role specifically in a research setting where the quality of data, transparency of data processing, and reproducibility are the key factors.

The conceptualisation of metadata conceptual framework of statistical metadata goes beyond the traditional descriptive metadata conceptual framework to capture detailed information on aspects such as data collection methods, data processing procedures, variable definition and data quality assessment (Ramírez et al., 2017). Such a wholesome strategy allows researchers to assess the adequacy of the datasets to their respective analytical requirements, and it offers the contextual information that allows the statistical output to be interpreted appropriately. Ensuring statistical metadata integration with business intelligence systems has gained great significance to organizations that aim to use their data resources strategically in making decisions.

The statistical metadata standards have been developed to respond to the increased complexity of the data processing chains and transparency requirement in statistical production. An example of this evolution is the Structured Data Transformation Language (SDTL), which offers machine-processable formats of data transformation commands found in statistical analysis packages (Alter, Ionescu, et al., 2020). SDTL allows to automatically capture provenance metadata using data transformation scripts and generates detailed documentation of all processing steps and variable derivations.

Realizing statistical metadata demands proper attention to technical requirements and user requirements. It has been shown that the productivity and effectiveness of data analysis processes can be greatly enhanced by means of enriching data warehousing systems with structured statistical metadata (Jiménez-Ramírez a et al., 2017). This is a step beyond basic data management enhancement to enable advanced analytics, predictive models and machine learning applications where exact knowledge of data characteristics and quality are needed.

More recent statistical metadata systems are starting to use semantic web technologies to achieve better interoperability and machine-readability. More advanced statistical classification and metadata management is possible through the use of Resource Description Framework (RDF) and Simple Knowledge Organization Systems (SKOS) (Hancock, 2020). Such technologies facilitate the update of old hierarchical classification systems to more dynamic and receptive methods of the organization and discovery of statistical data.

## REFERENCES


Adamu Ahmed, A., Usman Abdullahi, A., Ya'u Gital, A., & Yusuf Dutse, A. (2024). *View of The Role of Metadata in Promoting Explainability and Interoperability of AI-based Prediction Models*. Stecab.com. https://journals.stecab.com/jemr/article/view/153/101

Alter, G., Donakowski, D., Gager, J., Heus, P., Hunter, C., Ionescu, S., Iverson, J., H.V. Jagadish, Lagoze, C., Lyle, J., Mueller, A., Sigbjørn Revheim, Richardson, M. A., Risnes Ørnulf, Karunakara Seelam, Smith, D., Smith, T., Song, J., Vaidya, Y. J., & Ole Voldsater. (2020). Provenance metadata for statistical data: An introduction to Structured Data Transformation Language (SDTL). *IASSIST Quarterly*, *44*(4). https://doi.org/10.29173/iq983

Alter, G., Ionescu, S., Donakowski, D., Gager, J., Heus, P., Hunter, C., Iverson, J., Jagadish, H. V., Lagoze, C., Lyle, J., Mueller, A., Revheim, S., A, M., Richardson, Risnes, O., Seelam, K., Smith, D., Smith, T., Song, J., & Jaydeep Vaidya, Y. (2020). *View of Provenance metadata for statistical data: An introduction to Structured Data Transformation Language (SDTL)*. Iassistquarterly.com. https://iassistquarterly.com/index.php/iassist/article/view/983/962

Ballinger, L., Karl, B., & Chiu, A. (2017). Providing Quality Rights Metadata for Digital Collections Through RightsStatements.org. *Pennsylvania Libraries: Research & Practice*, *5*(2), 144–158. https://doi.org/10.5195/palrap.2017.157

Crandall, E. D., Toczydlowski, R. H., Liggins, L., Holmes, A. E., Ghoojaei, M., Gaither, M. R., Wham, B. E., Pritt, A. L., Noble, C., Anderson, T. J., Barton, R. L., Berg, J. T., Beskid, S. G., Delgado, A., Farrell, E., Himmelsbach, N., Queeno, S. R., Trinh, T., Weyand, C., & Bentley, A. (2022). Metadata preservation and stewardship for genomic data is possible, but must happen now. *BioRxiv (Cold Spring Harbor Laboratory)*. https://doi.org/10.1101/2022.09.12.507034

Formenton, D., & Souza Gracioso, L. de. (2022). *Vista do Padrões de metadados no arquivamento da Web | RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*. Unicamp.br. https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8666263/27830

García, R., Cediel, A., Teixidó, M., & Gil, R. (2023). A Review of Media Copyright Management using Blockchain Technologies from the Academic and Business Perspectives. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2307.16244

Glowacka-Musial, M. (2022a). Applying Topic Modeling for Automated Creation of Descriptive Metadata for Digital Collections. *Information Technology and Libraries*, *41*(2). https://doi.org/10.6017/ital.v41i2.13799

Glowacka-Musial, M. (2022b). *View of Applying Topic Modeling for Automated Creation of Descriptive Metadata for Digital Collections*. Corejournals.org. https://ital.corejournals.org/index.php/ital/article/view/13799/11237

Hancock, A. (2020). The use of metadata modelling for the modernisation of information management of statistical classifications. *Statistical Journal of the IAOS*, *36*(3), 797–806. https://doi.org/10.3233/sji-200665

Jiménez-Ramírez a, C., Edith Burke b, M., & Rodríguez-Flores, I. (2017). *Vista de Metadatos estadísticos en el descubrimiento de conocimiento | DYNA*. Unal.edu.co. https://revistas.unal.edu.co/index.php/dyna/article/view/61417/62461

Klaris, E. (2015). Getting to ROI: Rights metadata and the smart content life cycle. *Journal of Digital Media Management*, *3*(4), 342. https://doi.org/10.69554/jrhn3038

Lawson, M., Gropp, W., & Lofstead, J. (2022). EMPRESS: Accelerating Scientific Discovery through Descriptive Metadata Management. *ACM Transactions on Storage*, *18*(4), 1–49. https://doi.org/10.1145/3523698

Marjolein Steeman, & Schuurman, J. (2018). Preservation metadata for DPX files: A case study. *Journal of Digital Media Management*, *7*(1), 33–33. https://doi.org/10.69554/urog3892

Martorana, M., Kuhn, T., Stork, L., & Jacco van Ossenbruggen. (2024). Zero-Shot Topic Classification of Column Headers: Leveraging LLMs for Metadata Enrichment. *Studies on the Semantic Web*. https://doi.org/10.3233/ssw240006

Matusiak, K. K., Werling, S., Donovan, L., & Carlson, S. (2021). Rights Metadata in a Community Archive: Implementing Standardized Rights Statements. *Journal of Library Metadata*, 1–18. https://doi.org/10.1080/19386389.2021.1955586

Mosha, N. F., & Ngulube, P. (2023). Metadata Standard for Continuous Preservation, Discovery, and Reuse of Research Data in Repositories by Higher Education Institutions: A Systematic Review. *Information*, *14*(8), 427. https://doi.org/10.3390/info14080427

Pro-Metrics. (2024). *Iberoamerican Journal of Science Measurement and Communication*. Pro-Metrics.org. https://ijsmc.pro-metrics.org/index.php/i/article/view/91/79

Ramírez, C. J., Burke, M. E., & Flores, I. R. (2017). Statistical metadata in knowledge discovery. *DYNA*, *84*(202), 270–277. https://doi.org/10.15446/dyna.v84n202.61417

Signore, M., Scanu, M., & Brancato, G. (2015). Statistical Metadata: a Unified Approach to Management and Dissemination. *Journal of Official Statistics*, *31*(2), 325–347. https://doi.org/10.1515/jos-2015-0020

Steeman, M., & Schuurman, J. (2018). *Preservation metadata for DPX files: A case study*. Journal of Digital Media Management. https://hstalks.com/article/768/preservation-metadata-for-dpx-files-a-case-study/?business

Xie, J., Zhang, M., & Ma, Y. (2019). Using Format Migration and Preservation Metadata to Support Digital Preservation of Scientific Data. *Using Format Migration and Preservation Metadata to Support Digital Preservation of Scientific Data*. https://doi.org/10.1109/icsess47205.2019.9040683