
AI translation • View original & related papers at
chinaxiv.org/items/chinaxiv-202310.02231

Post-print Introduction to the National Standard “Unified Content Label Format Specification”

Authors: Yang Peng

Date: 2023-10-08T00:00:00+00:00

Abstract

, content author, content entity, dissemination path, content digital signature, content responsibility chain, and UCL package digital signature. The standard specifies the definition, property category, property length, property payload, and payload data type format information for each field.

Full Text

Introduction to the National Standard “Uniform Content Label Format Specification” (ChinaXiv Cooperative Journal)

Author: Yang Peng, Associate Professor and Deputy Director, Future Network Research Center, Southeast University

On December 29, 2017, the national standard of the People’s Republic of China, *Uniform Content Label Format Specification*, led by the National Chinese News Information Standardization Technical Committee, was formally approved and promulgated by the General Administration of Quality Supervision, Inspection and Quarantine and the Standardization Administration of China, with the standard number GB/T 35304-2017. The main drafting organizations included Southeast University, Xinhua News Agency, the State Administration of Press, Publication, Radio, Film and Television, Tsinghua University, Dayou Digital Resources Co., Ltd., and Peking University. The Uniform Content Label (UCL) is an original content metadata framework developed by Chinese scientists, and the formal promulgation of this national standard provides critical normative support for the promotion, popularization, and application of UCL technology, as well as for the standardized indexing, efficient sharing, and lawful governance of big content data.

1. Background of UCL National Standard Development

The development of the Uniform Content Label is closely related to China's major strategic needs. The Internet has become the primary carrier of cyberspace, yet its defects and shortcomings have become increasingly apparent. The Chinese government has called for "making good use of the Internet and managing it well," advocating for the creation of a national public cultural service system through "triple network convergence." In November 2013, the Decision on Major Issues Concerning Comprehensively Deepening Reforms pointed out the need to "promote the standardization and equalization of basic public cultural services" and to "build a modern public cultural service system" on this basis, while also "ensuring national network and information security." In February 2014, the first meeting of the Central Leading Group for Cybersecurity and Informatization emphasized that without cybersecurity there is no national security, and without informatization there is no modernization; cybersecurity and informatization are like two wings of a body and two wheels of a driving force, and China must transform from a network power to a network superpower. In December 2015, the Second World Internet Conference called for law-based cyberspace governance and enhanced international dialogue and cooperation to promote reform of the global Internet governance system and build a community of shared future in cyberspace. In April 2016, the Symposium on Cybersecurity and Informatization work noted that cyberspace is the common spiritual home of hundreds of millions of people, and that a clean and healthy cyberspace serves the people's interests. Cyberspace is not "outside the law," and law-based governance must be strengthened. The Uniform Content Label fully considers these strategic needs, adopts a content-driven advanced concept for a new design of content identification, and forms an innovative tripartite identification system for big content data that integrates production, consumption, and management, making it a new content primitive for cyberspace.

The UCL concept was first proposed around 2000 by Academician Li Youping of the Chinese Academy of Engineering, based on in-depth research into Internet information sharing methods and their deficiencies. The initial research findings on UCL were mainly published in journals such as *Chinese Engineering Science* and *Television Technology*. In 2002, *China Media Technology* published a series of dialogues between Academician Li Youping and Comrade Sun Baochuan from Xinhua News Agency about the post-World Wide Web era and conjectures regarding the "fifth media," providing an early and relatively systematic exposition of UCL and new information sharing networks. Since the concept was proposed, UCL's connotation and extension have been continuously enriched and improved through years of research and application practice by Academician Li Youping and numerous scientists, domain experts, and researchers.

Furthermore, UCL research is closely associated with the latest development trends in the media field. In June 2015, Apple launched Apple News at its Worldwide Developers Conference, providing iOS users with customized news content services that aggregate content from popular websites and newspapers

according to user needs. Facebook also introduced Instant Articles, an interactive media content creation and publishing tool, during the same period. In October 2015, the *New York Times*, to meet the challenge of Apple and Facebook entering the media industry, proposed developing “Particles Code” to achieve accumulative news by organizing news reports along timelines and knowledge points through searchable “news codes.” The Chinese government attaches great importance to the new trends of media convergence and transformation in the Internet era. The central government’s *Guiding Opinions on Promoting the Integrated Development of Traditional and Emerging Media* clearly states that traditional and emerging media should leverage their respective strengths and develop in an integrated manner, promoting deep integration in content, channels, platforms, operation, and management. The Uniform Content Label conforms to the development trends in traditional media, emerging media, and the Internet, focusing on supporting the aggregation, dissemination, management, and personalized services of content information according to the requirements of media convergence development, emphasizing the improvement of massive content aggregation and analysis capabilities, and establishing an active service mechanism of “content finding netizens.”

The Uniform Content Label was developed based on the Uniform Resource Locator (URL), the foundation of the Internet. While URL addressing (“Locator”) has greatly promoted Internet development, URLs typically only represent the location of information resources and cannot describe their rich semantics, leading to problems such as difficulty in finding and managing Internet information resources and disorderly distribution. Specifically, the main drawbacks of the Internet include: useful information is not easy to find, individual needs cannot be met; online content is difficult to govern effectively, with 良莠不齐 content and spam flooding; network guidance is seriously lacking, and public opinion guidance is difficult to implement. Although Tim Berners-Lee proposed the Semantic Web concept, attempting to make the Web an intelligent network that can automatically understand words, concepts, and their logical relationships, the Semantic Web has proven extremely difficult to implement. The Uniform Content Label addresses the fundamental causes of difficulties in finding, managing, and disorder in Internet content resources, taking into account three important roles in content sharing applications (readers, authors, and managers), and effectively compensates for the semantic and management deficiencies of URL.

2. Development Process of UCL National Standard

Considering that the Uniform Content Label had undergone more than ten years of research and development since 2000, and its core ideas and main structure had matured, Academician Li Youping began planning the standardization of UCL in early 2013 and led the Future Network Research Center team at Southeast University in extensive preparatory work. In November 2013, at the second director’s office meeting of the National Chinese News Information Standard-

ization Technical Committee, a clear requirement was put forward to “establish a ‘Uniform Content Label Standard Working Group’ to accelerate the formulation of the Uniform Content Label technical standard.” In February 2014, the Committee’s secretariat invited Academician Li Youping to serve as the working group consultant and formally established the “Uniform Content Label Standard Working Group” dedicated to the formulation and promotion of the Uniform Content Label technical standard. The working group was mainly composed of experts and scholars from institutions including Southeast University, Xinhua News Agency, the State Administration of Press, Publication, Radio, Film and Television, Tsinghua University, Dayou Digital Resources Co., Ltd., and Peking University.

After its establishment, the working group quickly began developing the core standard *Uniform Content Label (UCL) Format Specification*. In March 2014, the working group held its first meeting to discuss and review the report on UCL research progress and standard development, reaching consensus on the format of the code and property sections in UCL. On this basis, the working group compiled and summarized the main research results of previous UCL technology, organized multiple discussion meetings, conducted in-depth research on relevant technical standards at home and abroad, and after fully absorbing opinions and suggestions from all parties, collaborated on the draft. By early May 2014, the drafting of the *Uniform Content Label (UCL) Format Specification (Draft for Comments)* was completed. After extensive solicitation of comments and revisions, the draft was submitted for review and successfully passed the review meeting in August 2014. In January 2015, it was officially released as a Xinhua News Agency technical standard.

Based on the Xinhua News Agency technical standard *Uniform Content Label (UCL) Format Specification*, the Committee’s secretariat applied to the Standardization Administration of China for national standard project approval in March 2015, which was formally approved in July 2015. Over the following two years, the working group held multiple meetings and academician-expert consultation seminars around the compilation of the national standard, continuously incorporating new requirements such as media convergence development, cyberspace security, and Internet content governance. The standard text was innovatively revised and improved in terms of UCL description completeness and structural flexibility, completing the review of the recommended national standard, as well as the draft for comments, submission version, and final approval version. In May 2017, it was submitted to the Standardization Administration of China. On December 29, 2017, the *Uniform Content Label Format Specification* was officially promulgated as a national standard of the People’s Republic of China with the standard number GB/T 35304-2017, and the standard was formally implemented on April 1, 2018.

3. Main Content and Features of UCL National Standard

The national standard *Uniform Content Label Format Specification* scientifically and comprehensively defines the basic format of UCL data packets, including the names, meanings, and definition rules of various fields. The UCL data packet is mainly divided into two parts: the UCL Code section and the UCL Properties section. The UCL Code section stores code information related to content for rapid reader-oriented navigation, with main fields including version, media type, priority and strategy, content source, primary category, secondary category, topic, content copyright and type, security level code, language, timestamp, and serial number. The standard specifies the definition, position, length, value range, and data type format information for each field. The UCL Properties section stores semantic information and management information related to content, with main fields including the UCL Properties header, content title, content keywords, content abstract, content author, content entity, dissemination path, content digital signature, content responsibility chain, and UCL package digital signature. The standard specifies the definition, property category, property length, property payload, and payload data type format information for each field.

The national standard *Uniform Content Label Format Specification* targets the development goals of China's future information sharing networks, focusing on the thorny problems of difficult-to-find, difficult-to-manage, and disorderly content on the Internet. Guided by Internet thinking and big data thinking, it provides standard support for integrating rich-type, multi-channel, and wide-domain national information resources, forming an innovative content big data identification system with Chinese characteristics. The trend of big data and rapid Internet development has driven media transformation, evolving the Internet into a gathering place for "broad, large, refined, and deep" content big data. Achieving universal scientific access to content big data inevitably faces three challenges: how to make content sharing benefit all people, how to ensure secure and trustworthy content services, and how the state can govern content according to law. The Uniform Content Label national standard is a powerful tool to address these challenges. It can improve the aggregation and analysis capabilities of content big data, enhance the dissemination power and influence of content, establish a "content-centered" universal guidance mechanism for all people, and contribute to "accelerating the construction of a new pattern of public opinion guidance."

The national standard *Uniform Content Label Format Specification* focuses on and serves major strategic needs such as China's modern public cultural service system, cyberspace security, and media convergence development, filling a major gap in China's media technology field and original content metadata standards. The Uniform Content Label is applicable for indexing content resources primarily aimed at mass audiences in media, Internet, and related industries. Its application scope includes not only news media such as news websites, information portals, news agencies, radio, television, newspapers, and online media,

but also other applications such as emergency broadcasting, content aggregation, public opinion analysis, and copyright protection, and it has already been applied in relevant fields. Compared with existing various content metadata standards and content identification technical solutions, the Uniform Content Label features advanced concepts, simplicity and efficiency, and flexible structure. It can describe the pragmatic, semantic, and management information of content resources from multiple dimensions, allows tailoring and free extension according to application requirements, and can effectively support efficient aggregation and ubiquitous distribution of content big data, personalized active services, semantic analysis and knowledge extraction, certification and registration chain-of-custody management, traceability and lawful accountability, especially demonstrating significant advantages in content governance according to law and cyberspace security.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.