U.S. GOVERNMENT PRINTING OFFICE
Keeping America Informed | www.gpo.gov

Overview of GPO’s Authentication Program

For more than 150 years, the core mission of the U.S. Government Printing Office (GPO), Keeping American Informed, has remained unchanged. Since 1861 users have looked to GPO as a trusted source for Federal government information. The presence of the words “United States Government Printing Office” on a printed publication assures the public that the content inside expresses information as it was approved by a Federal government author. This assurance is strengthened by trust relationships established between all parties in the creation, production, and publication process. In the traditional printing environment, a printing specialist from Congress, a Federal agency, or U.S. Court contacts a GPO customer service specialist to submit a publication to be printed. The resulting printed government publication is made available to the public from the GPO’s online bookstore or through the Federal Depository Library Program (FDLP), a system of more than 1,200 libraries in the United States.

The Challenge

Today, GPO continues its role as a trusted source for disseminating official and authentic Federal government publications to the public. The traditional print production process has evolved, and the adoption of digital technology has changed the way publications are created, managed, and delivered to users of Federal government information. Electronic documents pose a unique challenge because they can be easily altered, which can lead to unauthorized versions of Federal government content. GPO must assure users that publications available from GPO websites are as authentic and official as the publications printed and disseminated by GPO, and that trust relationships exist between all parties in electronic transactions.

Current Approach

GPO’s role in the authentication of Federal government publications is to provide tools and evidence for users to determine the authenticity of content. Authentic content is content that is verified by GPO to be complete and unaltered when compared to the version approved or published by the content originator. Official content is content published by the Federal Government, at Government expense, or as required by law. As a result, official content is authentic content.

A primary tool facilitating authenticity is GPO’s Federal Digital System (FDsys) (www.fdsys.gov), which manages, authenticates, preserves, and provides access to Federal government information. The system has been developed as a comprehensive, systematic, and dynamic means for preserving any type of digital content, independent of specific hardware or software.

Maintaining content integrity means GPO must ensure that content has not been changed or destroyed without authorization, which is especially applicable to electronic information. Minor changes are difficult to detect, and if the integrity of electronic information is compromised it cannot be authentic. GPO has implemented four measures to assure users that the integrity and authenticity of content has not been compromised: digital signatures on PDF files, cryptographic hash values in metadata, evidence of the trusted digital repository through the FDsys archive and access platforms, and demonstration of chain of custody.

Digital Signatures on PDF Files

GPO first defined its strategy for content authentication in a 2005 white paper that described the agency's need to develop policies and create systems addressing the authentication of electronic government publications. In early 2008, GPO began applying digital signatures to PDF documents, starting with the President's Fiscal Year 2009 Budget. The digital signature confirms GPO as the trusted information disseminator and provides visible assurance that the electronic document has not been altered since it was signed and disseminated by GPO. Today, most collections in FDsys that are available in PDF format are digitally signed.

GPO uses a digital certificate to apply digital signatures to the PDF documents. This digital certificate is issued by a Certificate Authority (CA) upon receiving proof of identity. The path between the certificate and the CA allows the signature to be validated. When a PDF document digitally signed by GPO is opened, a blue ribbon appears, verifying that the document has not been modified since the signature was applied. Signature properties are then available if further information about the signature is needed.

Cryptographic Hash Values

GPO also uses cryptographic hash values to maintain content integrity. Upon submission to GPO's content repository in FDsys, a number is generated for each content file that is unique to the data inside it; this process uses a published algorithm, or cryptographic hash function. That string of numbers is called a cryptographic hash value or message digest and can be used to detect any changes to the content. Even the smallest changes to the file will generate a completely different hash value.

The cryptographic hash values are provided to users for every publicly-accessible file on FDsys in the PREMIS metadata file. Users can search for content on FDsys and, using the information provided in the metadata and publicly-available tools, independently validate that the content has not been altered since it was submitted to GPO.

Trusted Digital Repository

GPO demonstrates to the preservation and user communities the use of best practices for establishing the authenticity of content and maintaining integrity within GPO's FDsys. Through an external auditor, GPO is working towards certification of FDsys as a Trusted Repository with the assistance of Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC). TRAC is a report published by the Digital Repository Certification Project and created by the Research Libraries Group and the National Archives and Records Administration to help repositories objectively probe trustworthiness.

To fulfill requirements set forth by TRAC, GPO's FDsys security strictly controls access to electronic content in the repository. Users can submit new content and change the descriptive metadata about content, but it is not possible to open a file in the repository and actually change the content.

Chain of Custody

Finally, GPO provides chain of custody, or provenance, which is information documenting the history of the content information for digital libraries and archives. Chain of custody records the content source, the changes that have occurred since the content was created or acquired, and who has had custody of content. This gives users assurance as to the likely reliability of the content information. Chain of custody is especially important in helping users evaluate the authenticity of derivatives of content. These derivatives are created for two reasons: 1) to perform preservation processes and 2) to provide an alternative representation of content that is easier for users to use.

GPO collects and makes available to users information about each significant event in the lifecycle of content in the PREMIS metadata file. The event information includes what occurred, who triggered the event, what specific files the event affected, and the date and time of the event.

Next Steps

GPO is a trusted steward of authentic Federal government content. GPO's policies and technologies are developed around a user-centric approach to content authentication, where the agency provides tools to help users make determinations about the authenticity of a particular piece of content.

Future authentication capabilities GPO is evaluating include: technologies to enable bulk content integrity assurance of XML files; authentication of smaller, discrete units of information; and enabling digital signature and certification mechanisms for mobile devices.

GPO recognizes that as technology changes and the field of digital content authentication develops, requirements for policies and technologies will change. GPO strives to evolve and will continue to be flexible and adapt its practices to meet the needs of its users and content authentication best practices. Through continued collaboration with other Federal government agencies and content authentication experts, GPO expects to remain a leader in the field of content authentication.

For additional information about GPO's authentication program, see the 2011 authentication white paper *Authenticity of Electronic Federal Government Publications* http://www.gpo.gov/pdfs/authentication/authenticationwhitepaper2011.pdf.