

Tip Sheet

Unify and Normalize Siloed Data

Master Data Management Essentials for Small and Medium Sized Publishers



Master data management (MDM) is a technology-enabled discipline in which business and IT work together to ensure the uniformity, accuracy, stewardship, semantic consistency, and accountability of the enterprise's official shared master data assets."

Source: <https://www.gartner.com/en/information-technology/glossary/master-data-management-mdm>

What is Master Data Management?

Master data management (MDM) is a process to ensure uniformity, consistency, and centralized location for an organization's primary data assets.

In most cases a publishing business will work with information technology to centralize, manage, and present data to different parts of the business. Some publishers will work with outside suppliers to help them achieve this and may take a lightweight or in-depth approach to the process.

Why does MDM matter for smaller publishers?

In a small publishing business there can be challenges with important data being siloed. There can be problems with holding multiple accounts for the same customers, authors, and for membership organizations, their members. When data is messy and unconnected, it is difficult to gain insight into relationships with organizations and individuals with whom the publisher interacts. It is also difficult to provide holistic information to staff who would gain knowledge from the ability to access, compare, and analyze multiple pieces of data.

One of the solutions to these problems is to work through an MDM project, which establishes the concept of a 'single source of truth' in data, so that all data about organizations, authors and members is unified. Where data is not centralized, it is inevitable that different sources of data will diverge and change from one another over time. MDM addresses the process for transforming data into a central source, where data are uniquely identified, consistently classified, and quality checked. The central source can then supply the data back to systems (and people), which were previously siloed so that all remain in sync and data can be connected to one another via persistent identifiers (PIDs). This means no more duplicate accounts, appending missing data, or correcting inaccurate data.

What challenges are there?

For some small and medium-sized publishers like university presses and societies, although certainly not all, there can be resource challenges. There may not be a dedicated data team, or people with the technical skills to manipulate and analyze data. For some publishers, there may not be the systems and tools in place to work with data efficiently and accurately. In these circumstances the concept of MDM may seem overwhelming.

Essential Steps to introducing MDM?

First, take a survey of data, involve the teams using or producing data, and ask them to document the data. Next, ask these basic questions.

What data is held?

- What does the data describe? e.g., organizations, or people?

Where is the data?

- E.g., spreadsheets, a CRM system, a finance system, third party systems?

Who has access to the data?

- Would the membership team benefit from access to author data?
- Would the marketing team benefit from access to usage data?

How is data created and updated?

- Is it manual, is it coming from a third party, who controls the data, who maintains the data?

Look for data quality issues:

- Are there duplicates?
- Is data inconsistently described, e.g., free text entries rather than taxonomy terms?
- Is data out-of-date?
- Are there things missing?

What initiatives are being considered for the future and what data would be required to make those business decisions?

Work out the priority areas for the organization that will hold the biggest benefit, and at the same time will reduce manual work and produce efficiencies. With established goals, there are then steps that can improve data quality and consistency and maintain that status.

Apply persistent identifiers (PIDs) to data — this will enable the unification of accounts and map one set of data to another if it is concerning the same type of entity. Select a PID schema which already has unified and consistent data and uses descriptive taxonomies and ontologies to define and enrich data. By adopting that schema, it will enhance the data held, and correct inconsistencies. Regular refreshing against the PID provider's data will ensure data is maintained in a consistent manner.

Generate a data governance policy and ensure data entry and maintenance conformity. For example, apply taxonomies to data and prevent free text fields as much as possible. Implement required fields and use validation in systems wherever feasible.



Learn More

CCC's Ringgold Solutions can help you on your first steps toward MDM, not only with our PID schema and additional rich metadata, but also Ringgold Audit Service. CCC offers a range of services to meet the needs of publishers starting the MDM journey.

Harness the advantages of MDM

With data unified via PIDs and metadata schemas applied to it, and with the data stored in a central place for systems and people to interact with and consume, publishers can achieve the following benefits:

See all of the interactions with an organization, such as:

- Subscriptions
- Book purchases
- Open access agreements
- APC payments
- Author affiliations
- Member affiliations
- Access and Authentication
- Entitlements
- Usage
- Conference sponsors

See all of the interactions with people, such as:

- Authors
- Members
- Editors
- Conference submissions
- Conference attendees
- Continuing Professional Development enrollment
- Other membership programs
- Peer reviewers

Connected data makes it possible to:

- Assess and formulate open access agreements and by joining subscription and APC payment data.
- Attract new authors, peer reviewers, and members by comparing the overlaps and gaps across those groups.
- Ensure authors and users of content are automatically granted the correct entitlements to read or publish based on your agreements.
- Check for conflicts of interest in peer-review.
- Generate conferences and events on topics in locations where you know there is a hub of interest from publication data.

About CCC

A pioneer in voluntary collective licensing, CCC advances copyright, accelerates knowledge, and powers innovation. With expertise in copyright, data quality, data analytics, and FAIR data implementations, CCC and its subsidiary RightsDirect collaborate with stakeholders on innovative solutions to harness the power of data and AI.



Contact us:

- copyright.com/ringgold
- solutions@copyright.com