



Creating a Talent Identifier for the Entertainment Industry

August 28, 2019

Contents

Executive Overview.....	1
The Case for a Talent ID.....	3
Why identifiers are important (in general)	3
What should a Talent ID identify?	3
Why Talent ID is important	4
Why the time is right for Talent ID	4
Principal Use Cases for a Talent ID.....	6
Delivery of distribution metadata	6
Metadata acquisition.....	6
Portfolio licensing and acquisition	7
Residuals processing and communication with guilds	7
Marketing research and consumer recommendations	8
Characteristics of Successful ID Systems Applied to Talent ID	9
Support from a broad coalition of stakeholders	9
Close coordination with international standards bodies	9
Cooperative governance that includes commercial and non-commercial stakeholders	10
Tightly defined and minimal data model.....	10
Minimal personal information.....	10
International scope from the beginning.....	10
Interoperability.....	11
Open access.....	11
Low-cost implementation with a realistic financial plan	11
Trust	11
Additional technical requirements	12
Building the System	13
Collaboration and leveraging standards.....	13
Technical considerations around privacy protection	14
Technical models	14
Implementing the chosen technical model	15
Data model - domain metadata for film, TV, music, and sports	16
Matching and avoiding duplicate registrations	17
Populating the system	18

Seeding	18
Alternate IDs.....	19
Operating model	19
Updating records with new works and other new information	19
Tools for registrants	19
Timing of new registrations.....	20
Trusted registration by guilds, agencies, and other trusted parties	20
Appendix - Existing Systems	21
International standards	21
DOI (Digital Object Identifier).....	21
ISNI (International Standard Name Identifier)	21
Proprietary systems – IMDB & others	22
Open source system - Wikidata.....	22
Libraries and rights societies	23
VIAF (Virtual International Authority File).....	23
CISAC IPI (International Confederation of Societies of Authors and Composers).....	23
ORCID (Open Researcher and Contributor ID)	23
Film archives.....	23
Guilds and unions	24
Existing industry ID registry - EIDR (Entertainment ID Registry)	24

Executive Overview

An actor receives a residual check for an episode in which she did not appear. The wrong director is credited in a cable program guide. A cast page links talent to the wrong movie. Portions of a writer's social security number appear in a public document. These and other errors occur because there is no standardized method for identifying people in the media entertainment world. Instead, the industry relies on government IDs and manual processing of names and text. However, names are not good enough—many people have the same name, people have multiple names, names often have alternate spellings or misspellings across systems—and government IDs should be used only for secure transactions. The industry needs a talent identifier that can refer to people accurately and reliably across workflows.

Broad industry adoption of a new standard ID for people would deliver substantial benefits to the entertainment industry. It would decrease costs by reducing errors and increasing efficiency, lower privacy and security risks, improve analytics, and facilitate new means of promoting and finding content.

The need for a standardized, non-confidential Talent ID has never been more apparent. The industry has moved aggressively to adopt standard work identifiers such as EIDR (Entertainment ID Registry) and implement suites of automation initiatives around delivery of avails data, marketing metadata, package and manifest data, as well as others. Set against the progress occurring around automation of digital distribution, the lack of a standard Talent ID stands out as a problem the industry must address.

In parallel, new regulatory regimes such as GDPR (General Data Protection Regulation) in Europe and the CCPA (California Consumer Privacy Act) in California require the deployment of stronger protections for individual data, especially Personally Identifiable Information (PII) such as SSN's and other government IDs. Continued use of these types of identifiers for anything other than financial payments creates tangible risks—risks that a standardized non-confidential Talent ID could help address.

We also are now entering a time when emerging technologies such as machine learning and ontology-based inference engines will have the ability to take real advantage of automated collections of talent information and deliver meaningful improvements in the way that marketers interact with consumers and consumers interact with content.

Combined, these industry developments highlight the need today for a standardized Talent ID.

The industry now has broad experience with implementing new standard identifiers, having deployed EIDR successfully across many digital workflows. That experience offers insights and best practices that can be brought to bear, helping to define a reliable path to implementation of a standard Talent ID. As this white paper outlines, that path includes learning from existing ID systems, collaborating with international standards where available, bringing together a coalition of commercial and non-commercial stakeholders to drive adoption, tightly defining the scope of data collected (especially around PII), and building an open and interoperable system with components that enable commercial automation.

This white paper describes the need for a Talent ID, lays out the potential use cases for the industry, defines key requirements for building such a system, assesses the pros and cons of other ID systems, and proposes a path to building and deploying the new ID. Our recommended path is based on active collaboration with existing efforts like the International Standard Name Identifier (ISNI), the Digital Object Identifier (DOI), and EIDR. It includes assigning a unique identifier to creative talent associated with works (e.g., directors, composers, writers,

performers, athletes) and maintaining an open industry registry of talent names, the work associated with the talent, and cross-referenced alternate identifiers from other existing industry databases. The registry should be resolvable and optimized for automation in commercial use cases.

We look forward to deeper discussion of these issues and working with stakeholders across the industry to implement a solution.

The Case for a Talent ID

Machine-readable identifiers offer efficiency benefits and potential automation that cannot be achieved with the open text strings too often used for names and titles of works. The industry has made significant progress substituting standard title IDs (such as EIDR) for text strings. Now is the time to take similar steps to substitute a universal Talent ID for non-standard textual names.

Why identifiers are important (in general)

In media workflows today, there are two primary methods for matching people: string matching and identifiers.

String-matching finds a name like “Jon Smith” in one system, then looks for “Jon Smith” matches in another system, an approach which presents obvious problems. For example, which “Jon Smith” is correct if there are multiple? Is “Jon Smith” the same as “John Smith” or “John Smythe”? What if the person’s legal name is “Jonathan Smith” or “Jon K Smith”? What if the name is partially translated as in “Jan Smith” or “Johan Smith”? In addition, text strings can be formatted differently, e.g., first name followed by last name, last followed by first, comma-separated or not comma-separated. There are double-barrel names, initials, titles, suffixes, monikers and simple misspellings. On-screen names for an individual can vary from work to work. Naming conventions and script differences can diverge across partners, regions, and platforms.¹

To handle these problems, complex rules-based systems use additional data, like an address, to help disambiguate people. However, these systems are still not completely accurate, and despite expensive manual review, mismatches still persist.

Identifier matching is simple by comparison. In the case of Talent ID, each identifier uniquely identifies one person. If an ID in one system matches an ID in another system, that person is the same. There is no guessing, no need for other data to disambiguate, and no parsing of complex matching rules.

Despite the advantages of ID matching, the names of actors, directors and other creative talent are among the most commonly exchanged text strings in the digital distribution of movies and TV. However, media systems that exchange information based on standard identifiers, especially EIDR, are starting to address the deficiencies of string matching on a large scale. Just as EIDR was created to enable automated communication of title IDs, a standardized Talent ID can solve the same problems for names of people.

What should a Talent ID identify?

An identifier provides a standardized way of referring to something (i.e., a machine-readable reference number) and allows for unambiguous communication about what is being identified. Therefore, IDs are generally a prerequisite for enabling automated communication between machines without human intervention.

Identifiers are domain-specific. For example, EIDR identifies film and television works, while ISRC identifies musical recordings. Each is associated with its own domain-specific metadata. In any identifier system, it is crucially important to define clearly the exact set of things being identified. That set of things must be constrained

¹ For example, Jackie Chan, Chan Kong-sang, and 陳港生 all refer to the same person, at least in the context of films. A machine-readable identifier can include all of these as alternate presentations of the same name.

enough to narrow the range of metadata necessary for unambiguous identification, as well as complete enough to satisfy underlying business requirements for the identifier.

As a general matter, this white paper addresses identifiers for people in the entertainment industry. Our goal is to enable efficiency in the way that creative projects of talented people are described, packaged, discovered, and distributed across different workflows, platforms, and channels. This type of ID is described as a “Talent” ID because “talent” is less broad and more precise than alternatives such as “person” or “name”. In this white paper “Talent” means “someone who contributed to some end product in our industry”—directors, actors, producers, writers, musicians, composers, set dressers, costume designers, composers, VFX engineers, and so on. It includes athletes and other sports figures who entertain in athletic performances and often also contribute to TV, film, or video productions. Guilds and talent agencies keep close track of many of these talented contributors, as do some departments in the studios and other content producers.

This definition, although still somewhat loose, narrows our scope. It also tells us that a Talent ID must be built on metadata sufficient to identify, uniquely and unambiguously, contributors in the worlds of film, TV, music, sports, and books.

Why Talent ID is important

There are numerous potential benefits from the adoption of a standard Talent ID. First, Talent ID would streamline all types of metadata workflows involving people: (1) reducing errors in the presentation of data to consumers by replacing text strings with more reliable codes that can be resolved by the recipient, (2) cutting costs by decreasing the volume of manual processing and validation and matching around metadata deliveries, and (3) speeding the ingestion of metadata from vendors and partners involved in portfolio licensing and acquisition deals.

Second, Talent ID would integrate talent information across entertainment sectors with one ID identifying talent across film, TV, music, and sports; areas which overlap across numerous industry data and financial systems.

Third, Talent ID would eliminate a large privacy and security risk in the communication of rights and royalty data. SSN’s and other government IDs are used far too often to identify talent and eliminate confusion among content producers, guilds, unions, and agencies. A standardized Talent ID would eliminate the need for using tax IDs for precise identification and decrease privacy and security risks for the industry.

Fourth, broad adoption of Talent ID by metadata systems and vendors would enable more automated talent research for marketing and promotional purposes, making it easier to create and deliver on-the-fly packages of content to consumers around favorite performers, directors, composers, and other creative talent. It would also enable new and more accurate search and recommendation services for consumers.

Why the time is right for Talent ID

The increasing reliance on data for supply chain automation increases the urgency to adopt a consistent, cross system Talent ID. EIDR has paved the way for identifiers and helped create industry expertise around ID deployment and use. Cross-industry specifications have improved the efficiency of digital workflows around avails, marketing metadata, package and manifest data, and digital extras. GDPR and CCPA are generating increased interest in automated solutions for managing privacy and data security around people, including eliminating unnecessary exposure of PII such as SSN’s and other government IDs.

All of this is occurring while powerful new technologies (e.g., machine learning, AI, and ontology-based inference engines) are creating new opportunities to learn from automated collections of talent data and improve how marketers interact with consumers. The same technologies offer new capabilities for ingestion and matching of data streams to a universal ID, easing the costs of adoption and deployment of a new Talent ID.

Lastly, international standards around naming conventions are reaching a new level of maturity. The ISO standard for identifying people, the International Standard Name Identifier (ISNI), has made strides in implementing API-based systems and other capabilities that offer a strong foundation for a new entertainment industry Talent ID. Together with industry experience around EIDR, the international standards and operational know-how needed to deploy a Talent ID are largely in place, simplifying the path to implementation

Given the convergence of industry trends, legal requirements, new technologies, identifier expertise, and standards capabilities, the current media ecosystem is ripe for deployment of a standard Talent ID.

Principal Use Cases for a Talent ID

For many of the same reasons cited above, there also is no shortage of potential use cases for a Talent ID.

Delivery of distribution metadata

Accurate metadata is a fundamental requirement for all channels of content distribution. In an ever more automated distribution environment, accurate and actionable metadata delivers revenues by enabling accurate presentation of relevant product to consumers.

In the theatrical window, studios deliver marketing metadata to exhibitors, marketing partners, and app developers. In the hospitality window, in-flight and in-hotel service providers need metadata to present consumers with appropriate viewing choices. Both electronic and physical home entertainment retailers require metadata to create and link web and in-store sales opportunities. MVPD's require metadata to ingest and automate high volumes of pay-per-view and free-to-subscriber VOD content. Broadcast networks and channels, as well as linear and non-linear programming guide vendors, need metadata to inform consumers about content available for viewing. Finally, SVOD and D2C streaming services could not succeed without accurate metadata.

Through all distribution windows, content operations and marketing teams are in near constant communication with video sites, trailer distribution partners, artwork distribution partners, advertising platforms, press, fan sites and other film/TV marketing partners to provide information in the form of metadata to generate excitement, buzz, views, and revenues. Similarly, in all windows, search and recommendation services use metadata to identify likes and dislikes and connect consumers to content and talent that interests them; for example, Amazon X-Ray informs viewers about what or who appears in videos and where.

In every one of these interactions there is the potential for miscommunication and error, and that potential is increased by manual interactions and text string communications that cannot be automated sufficiently. Distributors spend time ingesting, normalizing, and fixing textual metadata from hundreds of content partners before presentation to consumers.

Reliance on text strings for people's names inevitably results in operations and marketing staff spending time responding to customer support requests, clarifying metadata questions, and fixing errors. A key way to reduce that support time, and increase the accuracy of data presented to consumers, is to automate as much as possible and replace text strings with precise machine-readable codes when reasonable to do so. A standard Talent ID shared by studios, distributors, and marketing partners offers the opportunity for increased automation, elimination of errors, and reduction in manual processing and customer support time.

Metadata acquisition

In addition to the challenges of metadata exchanged with distribution partners, studios, broadcasters, digital platforms, agencies, guilds, and service providers all face the challenge of ingesting metadata from multiple other sources. All media companies acquire analytics data to support marketing and measure performance. Audience measurement data, performance data, sales data, social media data, partner reports, and other data are delivered with title and talent metadata from dozens of metadata vendors, service providers, and partners. Data in different formats with different naming and other conventions must then be integrated into internal systems used for finance, marketing, rights, and title management.

Almost every use of data tied to talent requires text-string matching for talent names. The same manual processing and potential for error that burdens distribution metadata then burdens data warehousing operations ingesting data for analytics and performance measurement. The result is severely limited ability to automate key business tools, slowing business processes and the delivery of critical data to decision-makers, and restricting new business opportunities that require faster turnaround and less expensive, more streamlined data management.

Portfolio licensing and acquisition

The same issues add costs and slow the work of operations teams dealing with new portfolios; whether acquired by purchase, licensed from third parties, or shared with co-producers or co-distributors. In every instance the new portfolio of titles must be ingested into title management systems, along with information about the talent associated with those titles. The ingestion of titles themselves is made easier by a universal content ID such as EIDR, but talent data still requires a substantial degree of manual processing, slowing the work and increasing the costs of integrating any new portfolio of titles no matter how acquired. A standard Talent ID across all content producers and talent databases would go a long way toward easing that integration burden.

Residuals processing and communication with guilds

A standard Talent ID also offers the potential for more efficient—and importantly, more secure—processing of residuals and all the related guild and talent communications that accompany talent payments.

Today many Hollywood talent and guild communications rely on a combination of name and social security number (typically last four digits) for precise identification of talent—presumably, other tax IDs like SSN are used for communications globally. While sensitive tax IDs are required for payments, banking transactions, and communications with tax authorities, any broader use of tax IDs in correspondence, email, queries, reports, and other contexts exposes sensitive personal data unnecessarily and creates security and privacy compliance risks.

A standardized, nonconfidential Talent ID shared across guilds, agencies, talent, and content producers would reduce security and privacy risks and decrease costs of regulatory compliance in situations where extra security precautions are not required.

A Talent ID available for use globally would also allow residuals and finance departments to rely on one consistent and non-confidential ID for communicating with multiple guilds and talent organizations across territories, easing the need to track and reference many different IDs in global talent communications. It would allow for more thorough and precise tracking of individuals associated with increasingly global productions, including increased automation of activities that today require substantial manual reconciliation.

A standard Talent ID used by financial and residuals systems as well as title management and metadata systems also would enable more automated error checks on payment processing. Talent participation can vary by episode in television programs and can even vary by version and edit of a film or show. Linking of title IDs such as EIDR to a standard Talent ID would allow automated verification that correct talent are paid for each episode or version. Similarly, performance and sales data reported by external vendors or partners could be integrated more easily into financial systems to allow more accurate and efficient verification of financial data against external data sources, as well as cross-industry, cross-window, and cross-territory reconciliation for joint productions.

It would also be a valuable tool for speeding responses to payment queries and researching performance of films and shows when data is required from databases (such as, legal databases, IMDb, filmographies, or other internal databases) that are not financial and don't track performance data by tax ID.

The payment of talent requires multiple entities to communicate an array of timely and specific information about people involved in productions. A universal Talent ID has the capability to bring a higher level of efficiency, certainty and security to that process.

Marketing research and consumer recommendations

For marketing teams, a Talent ID would facilitate more automated talent research for promotional purposes, making it easier to create and deliver packages of content to consumers around favorite performers, directors, and other creative talent. It would also enable new and more accurate search and discovery services for consumers, including cross-platform search without the necessity of fuzzy text matching across platform data formats and search fields.² A registry that resolves alternate identifiers for talent also allows linking into other sources of information about a person, such as other people databases and social networks, such as Twitter and Instagram, as well as to works that a person has contributed to outside of the movie and television world, such as books or music. The result would be new opportunities for consumer recommendations and collections, especially when coupled with a standard identifier for the underlying works.

² Alternatively, when fuzzy text matching is required, a standard Talent ID with alternate names would enable better results. With multiple alternate spellings and formats referenced back to a single person, different name variants in a consumer search query could be resolved back to a single name and ID of record.

Characteristics of Successful ID Systems Applied to Talent ID

An identifier is a building block infrastructure for other systems, not a solution by itself. Infrastructure tends to have a slow uptake, both because changing existing processes is expensive and because adopters must trust all aspects of the system. That trust must cover the validity of the identifiers themselves, the coverage of the system (i.e., percentage of the talent pool identified), and the reliability of the system that provides it.

For validity the system must have a data model and sufficient metadata to guarantee assignment of unique identifiers and minimize the potential for duplicates and errors. Levels of required accuracy will vary based on different use cases, for example, distribution accuracy versus financial payment accuracy, but in the end the system must be designed to meet defined validity thresholds.

The initial coverage of the system must be broad enough to cover a large percentage of relevant talent, with easy mechanisms for qualified parties to add more. This will require early buy-in from some initial key stakeholders to seed the system.

Reliability is partly technical and partly financial. Scalable cloud-based architectures have made both aspects significantly easier to build and afford than a decade ago. But the system must be robust and responsive to adopters, who also must have confidence in the financial resilience of the new system.

Based on these general parameters and the review of existing systems (described in the Appendix), a Talent ID system for the entertainment industry ideally would satisfy at least the requirements set out in this section below.

Support from a broad coalition of stakeholders

By definition, a standard ID must have broad recognition and support in order to fulfill its purpose and achieve the coverage and financial viability necessary for success. To be truly successful, a media industry ID requires support from commercial participants in the ecosystem, industry associations representing relevant interest groups and international standards organizations. Commercial supporters of a standard Talent ID should include both producers and distributors of content associated with talent, metadata providers who organize information around the content, and other key service providers and application developers for the industry; especially organizations that work with talent directly such as talent agencies and payment service providers. Industry associations should include guilds and unions representing talent and relevant trade groups who provide technical support for producers and distributors.

Close coordination with international standards bodies

Experience with other identifier systems also has shown that working with existing industry and international standards can reduce duplication of effort and lower a significant barrier to adoption. Support from and collaboration with relevant international standards organizations such as ISO, DOI, and ISNI, as well as content archives and standards implementers (such as EIDR), will prove important to long-term success. That means partnering with other systems to seed the new Talent ID database and share IDs with those systems. It also means cross-referencing those IDs and linking to them for additional metadata that will not be contained in the Talent ID system directly, in turn making it easy for other systems to access and cross-reference a standard Talent ID. Finally, it potentially means collaborating directly on ID management and operations, as long as collaboration is consistent with the underlying requirements of an industry Talent ID.

Cooperative governance that includes commercial and non-commercial stakeholders

An organization of many stakeholders requires a governance structure that is responsive to those stakeholders, especially those that drive coverage and adoption. Support of industry-sponsored organizations such as MovieLabs, along with other non-profits and international standards groups, is important but not sufficient to drive adoption and viability. Direct industry participation in governance is a requirement, since an ID designed to serve commercial use cases requires strong industry support.

Tightly defined and minimal data model

An inverse requirement of broad stakeholder support is the need for a narrowly defined scope that all stakeholders can agree upon. The broader the scope, the more likely there will be disagreement among the stakeholders around issues like a potential overlap with commercial services, metadata coverage, proprietary intellectual property, governance, confidentiality, and personal and privacy rights of individuals identified in a Talent ID registry. To be successful, a Talent identifier should be an ID registry, not a general metadata service, focusing only on a narrow range of talent data necessary to assign a unique ID, ensure reliable deduplication, conduct meaningful searches, and link to other identifier systems. That limits the potential for disagreement and maximizes the chances of success.

The identifier also should have the right level of granularity for the purpose at hand. In this case, the purpose is to identify people, not personas, names, or companies that may be associated with people. As a practical matter, that purpose implies that people may be identified with multiple personas or alternate names, but each of those separate personas or alternate names is not itself given a separate identifier. While that assumption must be tested against real-world use cases, the going-in principle sets a stake in the ground that should reduce data model complexity and again maximize the chance of success.

Minimal personal information

A narrow data scope also minimizes the potential for running afoul of evolving privacy and security laws around personal information. The more PII that is collected to create talent identifiers, the more the chance that talent themselves or regulators will object to how the data is collected or maintained, and the greater the cost of maintaining security for that data.

For a Talent ID to be successful, it must be broadly available without confidentiality restrictions. The goal should be a public ID like EIDR and other DOI-based identifiers, modeled after a public library card catalog. The registry should have enough information about creative works associated with the person to identify them uniquely with an ID, but minimal other personal data in order to minimize the privacy and security risks that accompany PII.

If possible, registering an identifier should not require any PII. There will be some cases where including PII is necessary, but the system should focus on identifying individuals through the works associated with that person.

International scope from the beginning

The entertainment industry is global, as are the use cases envisioned for a Talent ID. Any new Talent ID database should address international considerations in the data model as well as infrastructure and stakeholders. Alternate names should be captured in native language as well as alternate language variants. Contributions to works and alternate IDs from commercial and non-commercial sources globally should be captured for all talent.

The registry itself should be built in close collaboration with international standards organizations and be prepared to interoperate with talent organizations and guilds globally. Ideally, initial stakeholders should include media and entertainment companies globally and their partners.

Interoperability

Tolkien notwithstanding, there is no one identifier to rule them all. Any new standard ID must rely on interconnections with other identifiers designed for different purposes or different scope. A new Talent ID for the entertainment industry should anticipate and plan for interoperability at its inception, with the ability to record alternate IDs for the same person as a core connection point with other ID systems. Interoperability across other identifier systems, with alternate IDs and links to those systems, will enable easier user access to metadata in other systems, broader collaboration, faster application development, and more information for deduplication.

Just as importantly, any new ID system should borrow learnings from other systems and avoid reinventing the wheel. The data model should map cleanly to the best international standards available and ideally should integrate, either tightly or loosely, to other international systems designed for compatible purposes.

Open access

Another key requirement is openness. For a Talent ID to be successful and broadly adopted, it must be accessible to both the coalition of stakeholders who create the ID system and the much broader group of companies and organizations who are their partners in the ecosystem. A Talent ID must be freely resolvable by the industry as a whole, not just the immediate stakeholders, including all vendors and partners who need to use the ID in commercial workflows. That requires public lookup access and the absence of use restrictions that would block the development of applications using the ID. It also requires a democratic approach to creation of new Talent IDs, allowing trusted participants across the ecosystem to create IDs as needed to serve workflow requirements.

Low-cost implementation with a realistic financial plan

Machine-readable IDs enable efficiencies but add short-term infrastructure costs. To achieve broad adoption and realize the broader cost savings, an ID should be provided through a low-cost model on a non-profit basis as a utility for all participants in the ecosystem. It must be launched with sufficient capitalization to see the new system through its start-up years and a realistic plan for achieving financial stability with reliable long-term funding from trusted stakeholders. The financial plan should not tax usage by stakeholders and the ecosystem, but encourage it, especially in the early years.

Trust

Registering new identifiers must be based on trusted parties and trusted data sources. Most registry systems are to some extent crowd-sourced, with the data often curated after the fact. Especially for identifiers that involve financial transactions or have higher security and trust requirements (e.g., communication about residuals payments for talent) it is important to curate the crowd as well.

That requirement implies a somewhat formal mechanism for granting and revoking trust to participants who must be relied upon to maintain the accuracy and reliability of the trusted data. When for any reason accuracy or reliability breaks down, there should be a fast and fair process for revoking and regranting trust, and fixing the mistakes that prompted the revocation, both in the database itself and in the offending party's processes.

Additional technical requirements

Other technical requirements for the deployment of a Talent ID system for the entertainment industry include:

- **Uniqueness** - The identifier must be unique within its context. Everything in that context has (or can have) an identifier.
- **API-based** - Applications need a well-defined way to create, match, modify, and resolve the identifiers. The system UI should be built on top of the API.
- **Multiple resolution formats** - It should be easy to provide new resolution formats for the identifier. Tying the system tightly to a specific technology would be a mistake. XML has been popular historically; JSON is popular now; CSV probably will be popular forever (except for software developers), so it's important to retain flexibility.
- **Scalable and resilient** - Although current technology requires a central system for registration and deduplication, read access should be scalable and globally distributed.
- **Fast registration response** - A registration should return quickly, whether the system returns a newly created ID or an existing, matched ID from the database. If quick turnaround is too challenging, there should be a system of provisional IDs tailored for workflow requirements.
- **Fast correction of mistakes** – The system needs a reliable and quick means of correcting mistakes when discovered. Allowing known mistakes to continue to populate the system has technical and operational consequences which can magnify over time.

Building the System

The requirements outlined above provide the basis for our recommendations on implementing a new Talent ID system. These recommendations include proposals for collaboration with existing systems, a technical model with some implementation guidelines, a streamlined data model, a deduplication methodology, an approach to populating the initial database, and some thoughts on operating the system.

Collaboration and leveraging standards

A key requirement identified above is collaboration with other industry and international standards where appropriate. Looking at the available collaboration options, a partnership with the ISNI identifier system, the DOI system, and EIDR appears to be the most viable path to implementation.

The mission of ISNI closely parallels the goals of an entertainment industry Talent ID. The ISNI data model offers learnings that will assist with the creation of a Talent ID. ISNI infrastructure also is technically close to what is required. The key area of divergence between ISNI and our proposed model is around governance. An industry-sponsored Talent ID requires a high level of industry engagement and buy-in, which implies a level of commercial control that ISNI today does not have.

We recommend engagement with ISNI on governance with the goal of adding more commercial and industry involvement into the ISNI governance system. In parallel, we recommend building a new entertainment Talent ID under the aegis of an independent industry-run non-profit with the ability to act as an ISNI Registration Agency. The industry non-profit will collaborate as closely as possible with ISNI while reserving its ability to conduct independent operations. For the near term, that means discussing with ISNI the increased involvement of commercial and industry-specific stakeholders, while maintaining a separate governance structure in order to ensure a high level of commercial stakeholder input and influence at the board level.

As a technical matter, the industry Talent ID can integrate with ISNI and share operations and infrastructure, while maintaining whatever independent database and operations are needed to achieve the required commercial objectives. The new Talent ID system could integrate deeply or lightly, depending on how technical requirements match up. The exact level of technical and governance integration should be a matter of discussion and engagement with ISNI.

In addition to working with ISNI, we recommend active collaboration with EIDR and DOI. EIDR has an existing industry governance structure and experience around all aspects of ID registration and deduplication. As a DOI RA, EIDR also offers a direct path to collaboration with DOI, which provides useful capabilities for the new system around data modeling, resolution formats, alternate ID linkages, and connections to other similarly structured DOI systems. For these reasons, we recommend that the proposed Talent ID be resolvable as a DOI.³

³ If the new Talent ID system is operated by EIDR, it will not require a new RA agreement with the DOI Foundation.

Technical considerations around privacy protection

The legal and regulatory framework around privacy and PII is a moving target with GDPR and pending legislation like CCPA in California. The simplest way to avoid legal concerns is to not collect or store any PII. Alas, any Talent ID system needs at least a name and a list of associated works for that person. This information is generally a matter of public record, but in some jurisdictions even public information must be removed from databases upon request.⁴ This problem cannot be made to disappear, but it is likely to be rare in our Talent ID domain, where individuals generally seek recognition and payment as contributors to creative works.

Including limited PII may be useful for deduplication, if not completely necessary in most cases. If PII is needed for deduplication, it could be hashed in the same way as suggested for private keys and used as a yes/no point of comparison for deduplication purposes.

As another practical approach to these issues, the system could achieve privacy compliance by passing all PII data to ISNI and its already GDPR-compliant systems and processes.

If it does become necessary to accept PII, correcting and removing data can be handled separately. Technically, correcting data is just a change to an entry in the database. Administratively, it requires only a defined process to follow. Removing data (such as birth date or gender) is more problematic, because the data could be added back through a future registration or other data update. The problem could be solved by field-locking, or by recording which fields are removed for privacy reasons and ensuring those fields are not repopulated.

This is a problem of governance, not technology. It does not need to be addressed before the system is started, but will be a consideration during implementation.

Technical models

Based on experience with EIDR and other identifier systems, there are some obvious technical options for implementing a new system:

1. Adopt and use an existing identifier system without creating a new database, new registration system, new API, or new deduplication system. This option could include changing the existing system to meet requirements, but it would not include building a new system, and would provide little control over how users interact with the system.
2. As a corollary to option 1, build an industry-specific front end (API, UI, tools, and processes) for an existing identifier system. This option ameliorates any issues with the underlying system and provides a better fit with other industry-specific systems such as EIDR, while enabling efficient reliance on the existing system's infrastructure and processes.
3. Build an independent centralized registry. The registry could be completely standalone or based on collaboration with one or more existing systems. This applies to both the technical infrastructure (as in the case of EIDR and DOI) and data records (as with local ISAN agencies and the international ISAN agency). Either way the new system would include a new database with its own API and UI, plus a deduplication mechanism. This option would provide much more control than the previous two, though at significant cost, and would not take advantage of existing infrastructure.

⁴ The Swedish Film Institute will not make any data available that contains information about a living person, including what films he or she acted in or directed.

Option 1 is the lowest cost path, but it runs a high risk of not being nimble enough to meet industry needs. Option 2 provides the benefits of proven standards and infrastructure combined with flexibility in how the system is presented to users. Option 3 gives the most control at the highest cost (including investment in a full deduplication system), but it also increases the risks of conflicts with other standards and decreases the likelihood of international acceptance. While option 3 might be the only realistic path if there were no suitable infrastructure on which to build Option 2, the existing ISNI system offers a viable platform for our purposes.⁵

Our recommended path therefore is option 2, using ISNI as the underlying system and EIDR and DOI as the models for the API and technical feature set.

Implementing the chosen technical model

Option 2 comes with known risks related to existing infrastructure, including the risk that performance or service guarantees will not be sufficient for API read access or registrations. To address those risks and other system requirements, the system as envisioned should have the following high-level components:

- An independent core registry that provides a full API, including registrations that are then passed on to ISNI. Different flavors of the API can be implemented as smart proxies (one for XML, one for flat files, one for JSON, etc.). The core registry manages the rules for valid data submission, roles and security, and communicating with ISNI.
- API proxies can simplify adding application-specific features (many of which, in our experience, are just about data formats), and can help reduce load.⁶
- ISNI communication. There are two approaches: (1) continually pull the subset of the ISNI database we care about (which may be hard to determine), or (2) rely on ISNI's SLA and keep all the data in ISNI, not locally. The second is technically easier but will require more discussions with ISNI about reliability and performance since it becomes a single point of failure. ISNI also allows RA's to take a full copy of the database, though this is a new feature and not widely used.
- Read-only registries or proxies. These are a subset of the central registry and are for resolutions only (and maybe queries). It should be possible to spin these up quickly to manage load and failover issues. These could be a built-in part of the API proxies, but that's mostly a code management question.
- DOI integration. There are two ways to do this: (1) create a new DOI (similar to the EIDR mechanism) using a Talent ID prefix, or (2) work with ISNI and the DOI foundation to turn the ISNI itself into a DOI (as is done with ISBN-A, the DOI form of an ISBN). Content negotiation through the DOI proxy works well with the API proxies.
- The system should be designed and built from the ground up as cloud-deployable. Cloud deployment technology comes from many vendors and cloud providers, so ideally the system will not be tied to a single cloud provider.

⁵ We also considered a fully federated model, but there are not currently enough API-based name identifier systems to make that feasible. Many of the user benefits of a federated system are also found in a centralized system with good coverage of external identifiers.

⁶ As a point of comparison, API Proxies would radically simplify some EIDR feature requests that in the current EIDR system would place a significant load on the core registry and complicate the software release cycle.

Any UI should be built on top of the APIs, rather than wired directly into the registry. The ISNI UI is fully general, but this system's UI can be tailored to the industry needs and does not have to include ISNI features and functions that exist only to support other industries.⁷

For the registry itself, there are two interesting platforms to investigate (along with 'build from scratch', of course).

- Node.js is a language well suited to building high-throughput servers.⁸ Significant and well-supported frameworks exist for integrating with cloud-based data stores and handling security, etc. It is open source with a large developer base and could be integrated with DOI through CNRI's Handle libraries.
- CORDRA, which is a follow-on to the system on which EIDR was built (both developed by the Corporation for National Research Initiatives – CNRI). It is used by another DOI RA and by several companies in the financial services industry. It is database agnostic and is designed to move easily across multiple kinds of data stores, with support for cloud-based deployment. It provides a fair bit of help for managing schemas and APIs and has pre-built DOI integration.

Data model - domain metadata for film, TV, music, and sports

The proposed system should keep only the metadata required to disambiguate one person from all other people. That means only (1) a set of names that can be searched, (2) a list of works associated with the person, (3) identifiers from other systems, and (4) optional dates of birth and death if required for disambiguation.

For the narrow scope of this project, the metadata required to disambiguate should be primarily the works to which a person contributed. For audiovisual works, EIDR is an easy choice of primary work identifier, with other available work identifiers (IMDB, ISAN, resolvable film archive IDs, Wikidata, etc.) also accepted and used where needed. For music, ISRC and ISWC are both well-managed identifiers. The CISAC IPI would serve as a useful check on these, but it is not publicly accessible, nor are ISRC and ISWC resolvable.

Alternate names can be quite complicated. For example, some name sources have one or more of: primary name, preferred name; translated name, transliterated name, Romanized name, westernized name; billed name, credited name; legal name, contractual name; real name (whatever that means), birth name; display name, sort name; pseudonym, alias, nickname, or moniker. Attempting to categorize and sort the full range of name complexity should be outside the scope of the system, which instead should allow entry of all alternate names without categorization.

Beyond the kinds of names, there is an underlying problem of representation of names. Representations can range from a simple string to a complex framework for family name, given names, suffixes, and monikers (i.e., frequently used additions such as "Scatman" Crothers).⁹ How far down this path to go should be driven by the expected use cases (e.g., how the system will be searched by commercial and organizational users), not the needs

⁷ For example, the detailed roles for singers and specific instruments for musicians.

⁸ MovieLabs has had good experience with it, having tried it with MongoDB, a proprietary document store, and an RDF-based triple store as data repositories.

⁹ A full form of representation is provided as an optional part of MovieLabs Common Metadata. See section 3.6.1 of https://movielabs.com/md/md/v2.7/Common_Metadata_v2.7.pdf.

of consumer-facing systems. Our going-in recommendation is that a name is just a string. This position is extreme, and likely will be challenged as deduplication and search requirements become clearer, but it is easier to start with less and add more than to start with more and take something away.

Beyond names, works, and alternate identifiers, the value of metadata for disambiguation drops off rapidly. Date of birth and death can be fuzzy; gender is complicated;¹⁰ place of birth and country of citizenship can be disputed or changeable. It is unwise to rely on these for disambiguation, although they may be helpful in resolving borderline cases. These additional categories do make the system more usable for a human sitting at a user interface or browsing through lists of people, but that advantage has to be weighed against data privacy issues (see below).

In sum, the initial proposal for metadata is:

- A set of names associated with the person. Different identities or personas associated with a person can be listed as alternate names but will not be registered as separate Talent IDs. Internationalized names or names in different alphabets also should be included as alternate names where available. Names should have an optional language tag.
- A list of works to which the person contributed, with title (or titles) and or/identifiers, and optionally the person's role in that creation.¹¹
- Identifiers for the person in other systems, using a DOI-standard representation of identifiers.
- Optionally, date of birth and death.

This metadata is sufficient for ISNI registration. The PII should be included only when necessary.

Individual users, of course, may be unhappy to find that the UI presents only very limited metadata around works, names, and other identifiers. Once again, the Talent identifier registry should not be confused with a metadata system, and the goal of the system should be unique identification of persons, not the provision of metadata for consumer-facing applications. To help at least partially with user disappointment, however, resolvable alternate identifiers should be clickable in the UI, as they are in EIDR. And although the minimal data may make the UI less useful generally, it should still be possible to search for names, works, role in those works (actor, director, and so on), and other identifiers.

Matching and avoiding duplicate registrations

Deduplication is an essential part of this system. It relies on the metadata submitted with potential new records plus the metadata already in the system. More metadata usually helps with deduplication, as long as the metadata is correct.

ISNI can deduplicate people with 2-4 associated works (i.e., titles, ideally with identifiers as well). Alternate IDs for the person help. Adding birth, death, or gender adds very little if enough associated works are submitted, but helps considerably for persons with few associated works, such as, an actor who has appeared in only one film. As an assist to deduplication, ISNI also supports identification of the role played by a person in a work, plus a limited

¹⁰ With respect to gender, the MovieLabs Creative Works Ontology and Common Metadata do offer a flexible model to cover most known cases.

¹¹ The role list in the MovieLabs Creative Works Ontology is fine-grained enough to start with. See 'Role Definitions and Mappings' at <https://movielabs.com/creative-works-ontology/>.

notion of ‘worked with’, so roles or known collaborators on a work can be listed as a way of distinguishing people with similar names or contributing works with similar names.

Once a database is populated, each new registration is compared against existing records to find matches. The simplest match relies on a person’s name or names, a list of associated works, and an optional list of other person identifiers. If *enough* of the works match a person already in the database, and the name meets some set of fuzzy matching criteria, then the person is matched. If alternate IDs also match, that adds weight to the decision. This type of simple match could be done through a traditional rules-based system—which is more or less how ISNI works today—or with some relatively simple machine learning (once the database is big enough to provide a good set of training data). [See Seeding discussion below.]

If no match is found and the metadata is sufficiently rich, the system should allocate a new ID and return it to the registrant. If the metadata give an ambiguous result, the system should allocate a provisional ID, which over time may be promoted from provisional or aliased to another ID. To meet the requirements for immediate response to an API call, the database should return an answer even when it is unsure. Bulk registration and matching can take more time, but there should be a way for the requester to check the status of the request.

Populating the system

SEEDING

IMDb has about 9 million total name identifiers in its system. Across a sample of about 200,000 films, IMDb identified approximately 660,000 people.¹² Wikidata identified about 500,000 people across the same sample of films. IMDb has more of the minor contributors, while Wikidata tends a little bit towards the Wikipedia notion of “notability”. However, Wikidata also provides cross-references to multiple other systems, including IMDb and VIAF, as well as name variants.

With appropriate licensing relationships established, seeding the new system with names from IMDb or other commercial sources could be a feasible early step. Seeding with Wikidata is also a possibility. Partnering with a talent agency like Creative Artists Agency would enable phased seeding of “top talent” as a test and, if successful, a more complete agency database. Combining talent agency data with another source would give a core of high-profile people with links to other systems. Film archives such as AFI and BFI could also provide large, high quality datasets.

Collaborating with talent guilds and unions, along with the residuals departments of studios, is also likely to be an early requirement. Since those databases rely heavily on PII, especially government IDs, it will be necessary to provided seeded data in a secure way with hashed or otherwise encrypted references to IDs or PII necessary to ensure the new Talent IDs can be imported back into existing systems.

Seeding of the database should be accounted for in the budget of the new system, with sufficient attention given to ensuring high quality of seeded data and preserving all alternate IDs necessary to return seeded data to registrants mapped to the newly created Talent IDs. If the initial data is not high quality, problems will show up

¹² This number is not based on full credits, however, so the same 200,000 films could have a far larger number of people if full cast and crew were included.

later on and can be hard to detect and correct in a systematic way. The process for seeding should be viewed as a starting point for the process used for bulk registrations during ongoing operations.

ALTERNATE IDS

External identifiers should connect this identifier system to other systems and workflows and allow users to find metadata in other systems. The second case is mostly served by public and “quasi-public” resolvable identifiers (Wikidata, IMDb, film archives). The same identifiers can serve to connect workflows (as is sometimes done with IMDb IDs for films), but specific workflows often use internal identifiers that have no meaning in the outside world, e.g., “so-and-so calls this thing CodeXYZ” or “this product offer is called ALID:XYZ”.

Internal opaque identifiers can be very useful for deduplication, if they come from a trusted source. There should be a way of marking certain types as more reliable than others, at least in the context of deduplication. Internal identifiers are especially important as a way of matching bulk registration requests back to the registrant’s data. However, some internal identifiers may not be suitable for external release, either due to private information contained in the identifier itself, or due to other security reasons. These internal identifiers can be sent to and from the registry suitably hashed.

Operating model

Collaborating with ISNI and EIDR should enable the new system to rely extensively on the operating knowledge of its partner organizations. That approach simplifies some of the ongoing operational challenges to running and maintaining the registry over time. However, there are still some key issues to address.

UPDATING RECORDS WITH NEW WORKS AND OTHER NEW INFORMATION

As people contribute to more creative works, information about those works should be added to the database. In ISNI, new metadata (identifiers and works) submitted with a registration request is added to an existing record if one is found, which means that the system becomes smarter as registrations increase.¹³ For example, ISNI currently (2019-05-29) lists two works for Gal Gadot: the song ‘A Place Called Slaughter Race’ and the movie ‘Ralph Breaks The Internet’, in which the song appears. If a new ISNI were requested for Gal Gadot using data from EIDR, no new ID would be issued, but the EIDR-supplied data (acted in *Wonder Woman*, *Batman vs Superman: Dawn of Justice*, etc.) would be added to the existing record.

This is an important means of maintaining updated information on associated works, which itself is important to accurate deduplication of records.

TOOLS FOR REGISTRANTS

Implementers are far more likely to adopt new infrastructure if there are processes, tools, and applications that ease the transition. Any new identifier system must provide a ready means for adopters to ingest identifiers into existing databases, strong matching tools that decrease the burden of integrating with the new system, and a network of service providers able to assist new adopters.

¹³ EIDR does not do this because the metadata is more complicated and in many cases harder to be sure about, but users add alternate IDs manually, which improves matching and utility.

Bulk registrants especially will need substantial help. Skimping on tools and registrant help runs a high risk of lowering the quality of the database. Ongoing help for new bulk registrants should be accounted for in the operating budget.

One of the lessons from EIDR is the need for specialized tools to help prepare data for registration. For this project, a simple but useful helper would be a tool that takes a film or TV show title, a date, and/or a director and returns a list of EIDR IDs that match. The goal would be to make it easy for anyone with talent data to find EIDR IDs for the works associated with the talent to enable easier registration of new Talent IDs. That would substantially lower the bar to usage by helping registrants get secure work identifiers. Developing simple interfaces to allow for the mapping and normalization of data fields from a proprietary system to the fields used internally for the Talent ID would allow data to be submitted directly. Another promising tool would be a machine learning system trained on the existing registry that could analyze new bulk requests before submission for registration.

Going even further, the new system could offer a machine learning tool in a deduplication-as-a-service model, helping database operators integrate the new Talent IDs into their internal databases in an efficient and proactive way.

Finally, the availability of applications and services that make use of the identifier is an important driver of adoption. A new Talent ID must lend itself readily to use in automation tools across all relevant entertainment industry workflows.

TIMING OF NEW REGISTRATIONS

Identifiers are most useful if issued early in a process or workflow; such as, when signing on to an agency, joining a guild, or starting to work on a film. In the last case, early talent registration will also encourage early work registration in EIDR.

As mentioned above, attempted ISNI registrations that find a match add any new data to the existing record, so a standard practice of continual and early Talent ID registration of all people associated with a new film or show would improve the quality of the database, even for talent already registered.

TRUSTED REGISTRATION BY GUILDS, AGENCIES, AND OTHER TRUSTED PARTIES

The minimum metadata required for a new Talent ID in ISNI usually includes one or more associated works. However, there will be occasions in commercial workflows when a Talent ID is required for a new contributor without examples of completed work to reference; for example, writers or actors who are new members of a guild or union, but whose work is not yet published or commercially released. In those instances, if the registration is requested by a trusted party such as a guild or union or talent agency, the system will need to issue a new Talent ID for workflow purposes despite the lack of metadata.

ISNI has the notion of a 'trusted registrant', with data from that registrant viewed as authoritative. The new system will need to work with ISNI to apply this notion to data from trusted registrants to enable new registrations at the time required for commercial workflows.

Appendix - Existing Systems

The idea of a Talent ID for entertainment is not itself new. Many ID systems are in productive use today, and any new system should not be expected to replace all existing systems.¹⁴ Rather, a new system should recognize the reality of multiple IDs and leverage existing systems where possible, both by borrowing the best features of those systems and providing mappings between alternate IDs. A state-of-the-art Talent ID system, cross-referenced with existing systems, would provide the optimal solution for the largest number of ecosystem participants. Existing systems could continue to use custom identifiers for specific purposes, while enabling the use of a standard ID to ensure talent references are accurate when data is exchanged across systems.

This appendix summarizes some of the best features and shortcomings of existing systems.

International standards

DOI (DIGITAL OBJECT IDENTIFIER)

The DOI system provides the underlying technical and philosophical underpinnings for EIDR. Its model for people metadata is compatible with the requirements outlined above.

DOI has strong technical support for communicating alternate identifiers to applications and already supports a network of linked identifiers. DOI also has a strong commitment to multiple resolution formats, providing a standard way of redirecting requests to alternate resolution endpoints, for example a web page, XML, JSON, or a rights statement. The underlying resolution mechanism is inherently cloud-based and can be used in a minimal format (as with EIDR) or exploited more fully.

DOI provides these services to its own Registration Agencies (RA's) and also to other ISO identifiers. For example, ISBN-A is the resolvable DOI form of a standard ISBN. DOI also offers a path to interoperability and collaboration with other global standards.

For these reasons, making a new Talent ID resolvable as a DOI is an option that the new system should consider strongly.

ISNI (INTERNATIONAL STANDARD NAME IDENTIFIER)

ISNI is the ISO standard for name identifiers. The title is deliberately fuzzy—some ISNI records identify names and some identify people.

Technically, ISNI is the closest to meeting the requirements listed above for a new Talent ID system. It has a publicly disclosed methodology for deduplication, based largely on works associated with a person, while allowing the consideration of other information if necessary. ISNI has a workable API that is still evolving. ISNI provides a system for bulk matching/registration via XML or spreadsheet. It includes a concept of 'trusted registrants' that could be leveraged by key participants in a new Talent ID system.

¹⁴ Experience has shown that users should not be expected to implement a wholesale replacement of an existing identifier with a new one. Linking between multiple existing identifiers, both internal and external, will continue as the norm despite the adoption of new international standards.

Importantly, ISNI is based fundamentally on the use of alternate identifiers, with the goal of serving as an open cross-reference service that links other identifier systems. ISNI also has an open database that is publicly resolvable with respect to information that can be made available without violation of privacy or security.

ISNI is strongest in the library and publishing world, though with strong participation from the music industry and active collaboration with YouTube, SoundExchange and others on music industry people. In aggregate, ISNI has approximately 10 million people records already registered.

Similar to other ISO standards organizations, ISNI functions through a network of RA's focused on particular sectors. There currently is no ISNI registration agency focused on the film, television, or sports world. However, as a side effect of work in other sectors, ISNI does have a fair number of records for film and TV contributors.

Governance is by a board of mostly libraries and publisher societies, including CISAC and OCLC. There is one commercial entity on the ISNI board today, although the board has indicated a willingness to accommodate more industry participation.

For all of these reasons, ISNI is the most obvious international standard upon which to base a new Talent ID system.

Proprietary systems – IMDB & others

There are several online databases with extensive cast and crew lists, often with an internal identifier. The internal identifiers vary in use externally; for example, the IMDb *tt* code is used extensively to identify film and TV works, but the corresponding person identifier, the *nm* code, is less commonly used, though some external systems (ISNI, Wikidata) reference it.

Often metadata is crowd-sourced, with a UI for adding new metadata, but details of data validation and deduplication systems are not public. Requests for changes and fixes take variable amounts of time to resolve. The crowd-sourcing feature makes managing sources and reliability challenging, especially for use cases that require high levels of trust. And like other online film and television databases, there is often only one name associated with each person, plus some inconsistencies in formatting of things like titles, initials, first/last, etc.

It is also questionable whether API's are setup for commercial-scale transactions involving a constant flow of requests to create, match, modify, and resolve identifiers. API's and the registries are typically proprietary, with licensing rules that are not guaranteed to meet requirements for free and open access.

All that said, sites like IMDb have large volumes of useful data and could prove valuable resources for associating names and people with works, as well as providing important external identifiers for cross-referencing. We do not see them as suitable platforms on which to build an industry Talent ID registry, but it may be worth exploring partnerships to seed and populate the registry over time and provide registry users the option to obtain deeper metadata from proprietary sources.

Open source system - Wikidata

Wikidata is also a crowd-sourced system, with all of the pluses and minuses that entails. It provides a SPARQL API, which is a powerful tool, but one that requires specialized knowledge to use.

Since its main focus is providing infrastructure for linked data applications, Wikidata references a wide variety of alternate identifiers, including IMDb and VIAF.

As with IMDb, it is not a suitable platform on which to build an industry Talent ID, but it is an extremely useful source of work/person associations and other identifiers, as demonstrated with both EIDR and the MovieLabs Creative Works Ontology.

Libraries and rights societies

VIAF (VIRTUAL INTERNATIONAL AUTHORITY FILE)

VIAF is run by libraries for libraries and library users. It aggregates multiple authority files (individual library data records) into a common searchable format. Its members are primarily academic and national libraries.

It covers works and people. Unsurprisingly, coverage for books and authors is better than coverage for audiovisual works and their contributors.

Many VIAF members also contribute to ISNI. It is probably best to think of VIAF as a source of opportunistic data, as it shows up in Wikidata or ISNI, rather than a primary starting point for a new industry Talent ID.

CISAC IPI (INTERNATIONAL CONFEDERATION OF SOCIETIES OF AUTHORS AND COMPOSERS)

The Interested Party Information (IPI) system is a semi-closed system that manages relationships between contributors, musical works, and music performances. Access is for CISAC members only.

The system works well in its context (primarily managing rights-related distributions), but has minimal coverage of non-musical talent, and access is tightly controlled.

CISAC is also a major supporter of ISNI. Although CISAC has not yet contributed IPI data to ISNI, IPI might at some point be a valuable indirect source of data through the relationship with ISNI.

ORCID (OPEN RESEARCHER AND CONTRIBUTOR ID)

ORCID is an example of another industry (scholarly publishing) solving the identifier problem for its own purposes. It relies on self-registration and is used extensively for the industry of scholarship (grant effectiveness, influence ratings, etc.) Scholarship analytics are different from commercial talent analytics, but several companies have built analytics applications using ORCID and CrossRef (the largest DOI registry for scholarly publications).

The ORCID model for populating IDs and its governance are specific to its uses and user community, but a new Talent ID system could learn from ORCID's success in enabling use of the ID in applications.

Film archives

Most national film and television archives track people as well as films and television shows. Film archives provide high quality data relating people to works. The coverage is best for works in each archive's national remit.

Access to the data varies by institution:

- The British Film Institute has a beta version of an API.
- The American Film Institute is working on API access to its data.
- The Swedish Film Institute has an API, but with restricted access during a GDPR review.
- The Danish Film Institute makes metadata available in XML over http.

Archives can be a valuable source of seeding data, and all of these archives are extensively cross-referenced to EIDR, either because the archives are EIDR members (BFI, AFI), or through Wikidata and side experiments (SFI, DFI).

AFI and BFI have expressed a strong interest in engaging with an industry-wide Talent ID system, and both have strong engagements with ISNI.

Guilds and unions

Professional organizations have internal identifiers based on the PII of the organization's members. The membership databases include information about works associated with members, but those databases are not broadly available or resolvable by the industry as a whole; nor are they supported by public API's. For a number of reasons, therefore, guild and union identifiers are not suitable for wider public use. However, professional organizations are an important stakeholder group for any industry Talent ID. Participation by guilds and unions and linking of professional identifiers to an industry Talent ID are likely prerequisites to success of a new system.¹⁵

Any new Talent ID system will need to focus advance planning and effort on enabling easy mapping and ingestion of the new IDs into existing professional databases. That effort likely will vary by organization, but best practices and strong matching tools could enable a reduction in technical effort as new organizations join the system. A key goal for a new system should be to collaborate early on with one or more guilds or unions both as a seeding source and as a way of understanding issues specific to this sector.

Existing industry ID registry - EIDR (Entertainment ID Registry)

EIDR does not identify people, but the EIDR community has a good understanding of the infrastructure needed to run a linked identifier system and interaction with other identifier systems. EIDR is also a good source of name/work pairs, since most EIDR records have at least a director or some actors (top-billed talent in most cases). Those records also would benefit from replacing the name text strings with a standard industry Talent ID.

EIDR is well known in the industry, and its API provides a good model for dealing with alternate IDs and multiple resolution formats (both based on DOI). It already has an ACL-based permissions model, which is necessary for dealing with varying levels of trust. The new system should adopt these models wherever possible, extending them as necessary.

EIDR is a well-placed candidate for operation of such a system, either independently or in collaboration with one or more international standards organizations.

¹⁵ Significantly, the ISNI system views a guild membership ID as a strong basis for issuance of an ISNI ID.