We are now Refinitiv, formerly the Financial and Risk business of Thomson Reuters. We've set a bold course for the future – both ours and yours – and are introducing our new brand to the world.

As our brand migration will be gradual, you will see traces of our past through documentation, videos, and digital platforms.

Thank you for joining us on our brand journey.

REFINITIV™

OPEN PERMID:
AN OVERVIEW OF HOW IT WORKS AND HOW IT IS EXPECTED TO DEVELOP.

BOB BAILEY, VP, CHIEF INFORMATION ARCHITECT, THOMSON REUTERS
TONI MCDERMENT, INFORMATION ARCHITECT, THOMSON REUTERS

JUNE 2015: DOCUMENT VERSION 1.0

THOMSON REUTERS™

CONTENTS

INTRODUCTION: THE OPEN PERMID MODEL OF IDENTITY

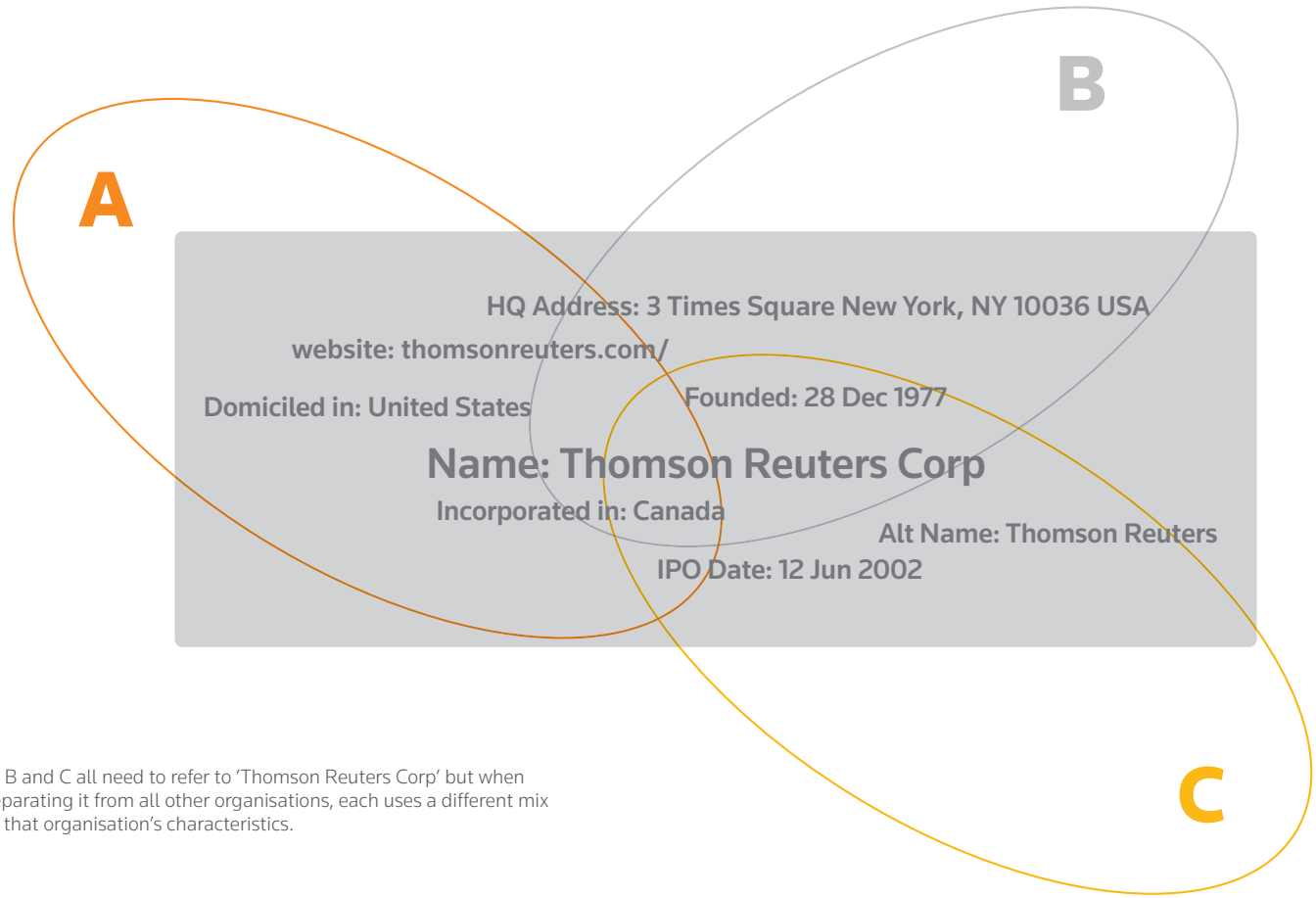
As data is exchanged and shared in an increasingly widespread fashion, the need grows for a means by which all parties involved can communicate identity. Identity is information intended to uniquely pick out an individual, salient thing in the real world. This information is captured as a set of defining characteristics; properties of the object most commonly used to differentiate it, but which do not require highly committed forms of agreement in advance of their use.

In human and machine communication, signifying the object of description or reference involves sharing an identifier that acts as a proxy for those defining characteristics. We 'dub' objects with names, knowing that when the name is used, we picture in our minds the characteristics by which we know that object. Machines give objects unique identifiers in order that the data recorded about the object may be retrieved, compared or updated.

Machine use of identifiers in communication has historically required all participating devices to agree on the selection, nature, and form of the characteristics used to uniquely distinguish the members of a given object type. For example, machines exchanging data on people need to have agreed upon the common characteristics of people that identify individuals as unique in the population concerned. (In this way of working, identity might be said to be 'tightly coupled' between the participants.)

The scale of the Web makes such agreement difficult to achieve and even more difficult to maintain, as new participants become independently involved, the population of individuals described increases, and the outlier set – members of the population who cannot be separated by the chosen characteristics – grows.

The solution to this problem is to adopt 'loose coupling' of identity. People do this all the time in conversation. When discussing a specific person, company, or any other object by name, there is no expectation of a previously agreed-upon set of characteristics that corresponds to the name.

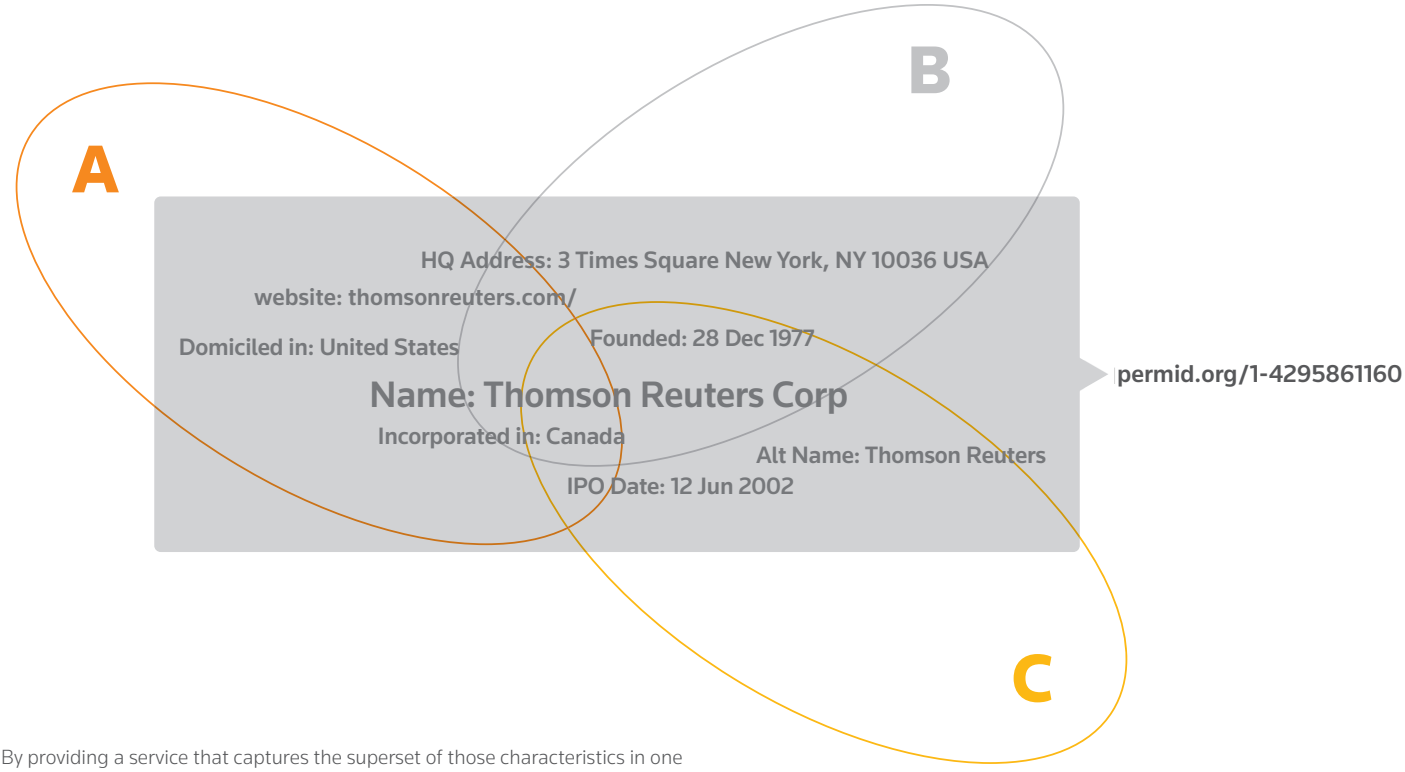


A, B and C all need to refer to 'Thomson Reuters Corp' but when separating it from all other organisations, each uses a different mix of that organisation's characteristics.

Instead we draw on our own experience. Provided each of us converts the identifier (name) used in communication to a set of characteristics that is unique to each of us and overlaps enough between us, there is a high probability that we are discussing the same person.

This 'loosely coupled' scheme of identity can be improved if the participants refer to a common lookup and synchronisation point. By pooling together a wider selection of characteristics by which individuals might uniquely identify the members of a set, and by giving each member of the set an opaque identifier[1] , participants can exchange the identifier and each use their preferred subset of characteristics to convert the identifier back into what it means to them. The maintenance by a shared authority of a managed set of characteristics for each object significantly raises the accuracy and precision of the method.

Furthermore, comparing the identifier for equality is the same as comparing the understanding of identity; if we both have the same identity for an object we know we are referring to the same object.



By providing a service that captures the superset of those characteristics in one place with a unique, immutable identifier, all parties can use that identifier as their reference point, each knowing that they can resolve that identifier to and from the characteristics they prefer to use.

1 It is important that the identifiers are opaque since if they were interpretable standalone, different participants might infer them to have different meanings or for such inference to differ from the characteristics held centrally. It is only by reference to the central registry that the correct meanings can be derived.

The operation of such a model might be as shown below:

1. An authority collects information about objects that are interesting to the community and how they are known and differentiated across the community. It does this using specialised understanding of the community and of the objects in question.
2. This information is used by the authority to build a registry of known objects.
3. Each object's registry entry is assigned an immutable Open PermID.
4. Users may perform two-way resolution of the information and the Open PermID according to whatever subset of the registry entry they are familiar with.
5. Consequently the Open PermID acts as a 'lingua franca' for identity: it can be ascertained from and converted to whatever subset of the registry data any of the users are familiar with. It can be used as an anchor for communicated information or as a consistent identity for workflow integration.
6. By contributing to the range of data inputs used by the authority, individual participants may improve the precision and range of information held in the registry.

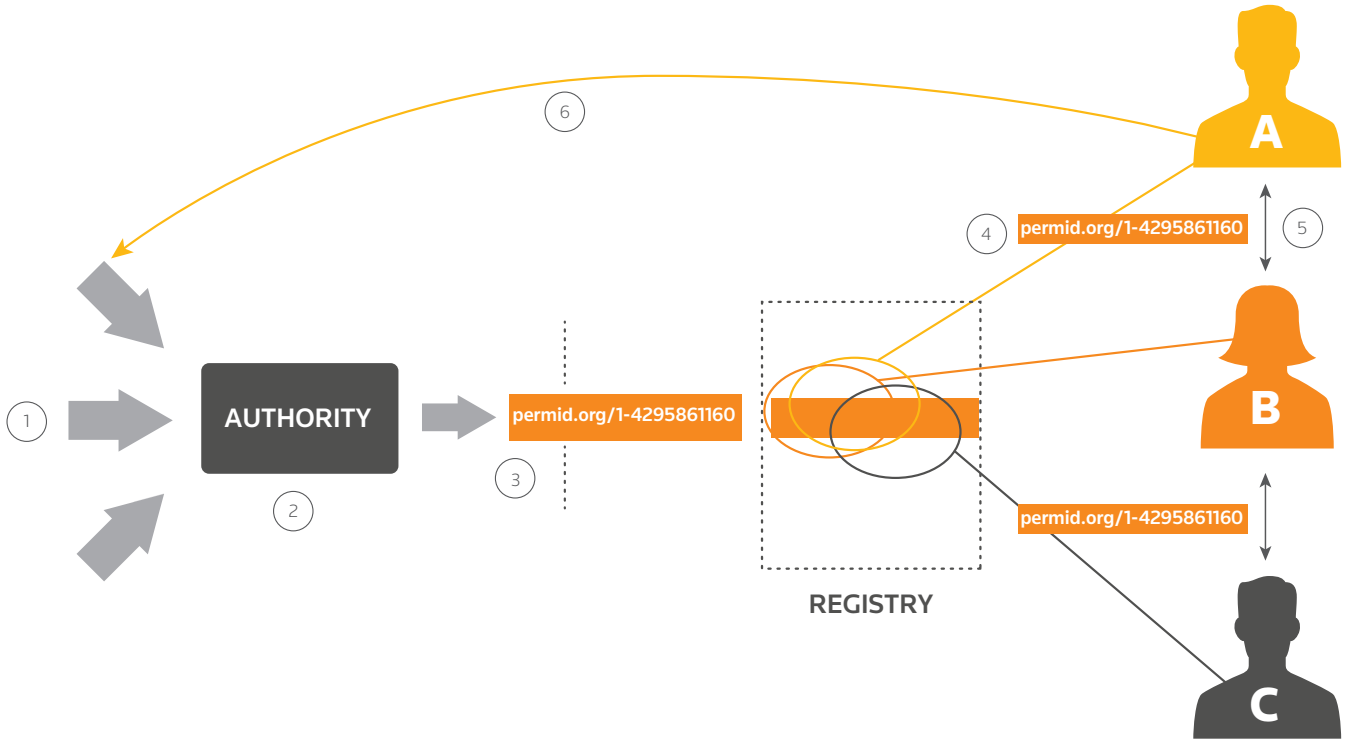
For example:

1. The authority sources data from major and minor providers concerning commercial organizations.
2. The authority disambiguates and matches/organises that data into a registry of unique organizations, each comprising multiple characteristics taken from across the sources.
3. Each organisation so represented in the registry is assigned an Open PermID.
4. User A searches the registry for an organization that has the characteristics:
   a. Name: Thomson Reuters Corp.
   b. Address: 3 Times Square,
      New York, NY, 10036, United States

   The Registry returns three 'hits'. All the characteristics for each entry are returned. From the additional data, User A selects the second hit as the right one. From this, he gets the **permid.org/1-4295861160**.
5. User A communicates some specific information concerning **permid.org/1-4295861160** to User B.
6. User B dereferences **permid.org/1-4295861160** using the registry and sees all of the characteristics for that entry.
7. From that, User B can see that she knows the object as the organization:
   a. Alternate Name: Thomson Reuters
   b. Incorporated in: Canada

This facility has the advantages, in that:

- No fixed scheme of identity has to be agreed by all participants in advance, yet the set of defining characteristics is understood by all.
- The facility decouples individual community participants. Each acts independently.
- Each participant has a lower cost, yet more precise, means of communicating identity (over meshed conversion, e.g., using mapping tables).
- The same facility can be extended to many sorts of objects and can be improved by participant contribution on an optional basis.



THE OPEN PERMID

The Thomson Reuters Open PermID is an opaque identifier that corresponds to entries managed in a federated registry of objects.

An Open PermID uniquely represents the defining characteristics of one specific object in the real world. Assignment of an Open PermID to a specific set of defining characteristics represents capture of the existence of that object.

The Open PermID facility is a mechanism by which groups who share a professional, commercially-based interest in the world can agree upon, pool, and share identity of the objects they need to describe and communicate. A shared facility ensures a stronger consensus and reduces overall costs. The Open PermID facility is designed to ensure that:

- Sufficient defining characteristics are maintained to enable common but isolated agreement on an object of description or communication.
- All Open PermIDs can be easily and predictably converted to the defining characteristics they represent.
- All defining characteristics can be searched as one universal set when looking for the appropriate Open PermID with which to label an object.
- Only one Open PermID corresponds to each identity (no duplication).
- Any given identity is seen by all as having a common state. All changes in its state are synchronized with all interested users as quickly as possible.
- Open PermIDs and identities share common rules and can be consistently used irrespective of the object identified or the authority responsible for maintenance.
- All Open PermIDs are persistent and the corresponding identities are held forever (no reuse, no deletion).
- Authorities may be any authorized party (they do not have to be operated by Thomson Reuters). 'Authorization' is by common agreement amongst the users of the Open PermID facility.
- Anyone using an Open PermID-enabled identity should provide feedback to the authority responsible for it, in order to improve its accuracy or provision.

These Open PermIDs are intended for machine use, not for human readability. Their purpose is to make machine communication of identity more precise and efficient. It is expected that they will be converted to human readable form before display to people, using the facilities described below.

Thomson Reuters will make available services for the lookup, reconciliation, dereferencing and synchronisation of Open PermIDs with the federated registry. These services are provided to customers using our products and on an open basis to the world at large.

The Open PermID is created as a URI, with the following syntax: **permid.org/xxx-xxxxxxxxxx**

There is a numeric element to the Open PermID, which is referred to internally as a PermID and conforms to the same rules as the Open PermID. This is what will be exposed in our existing product models.

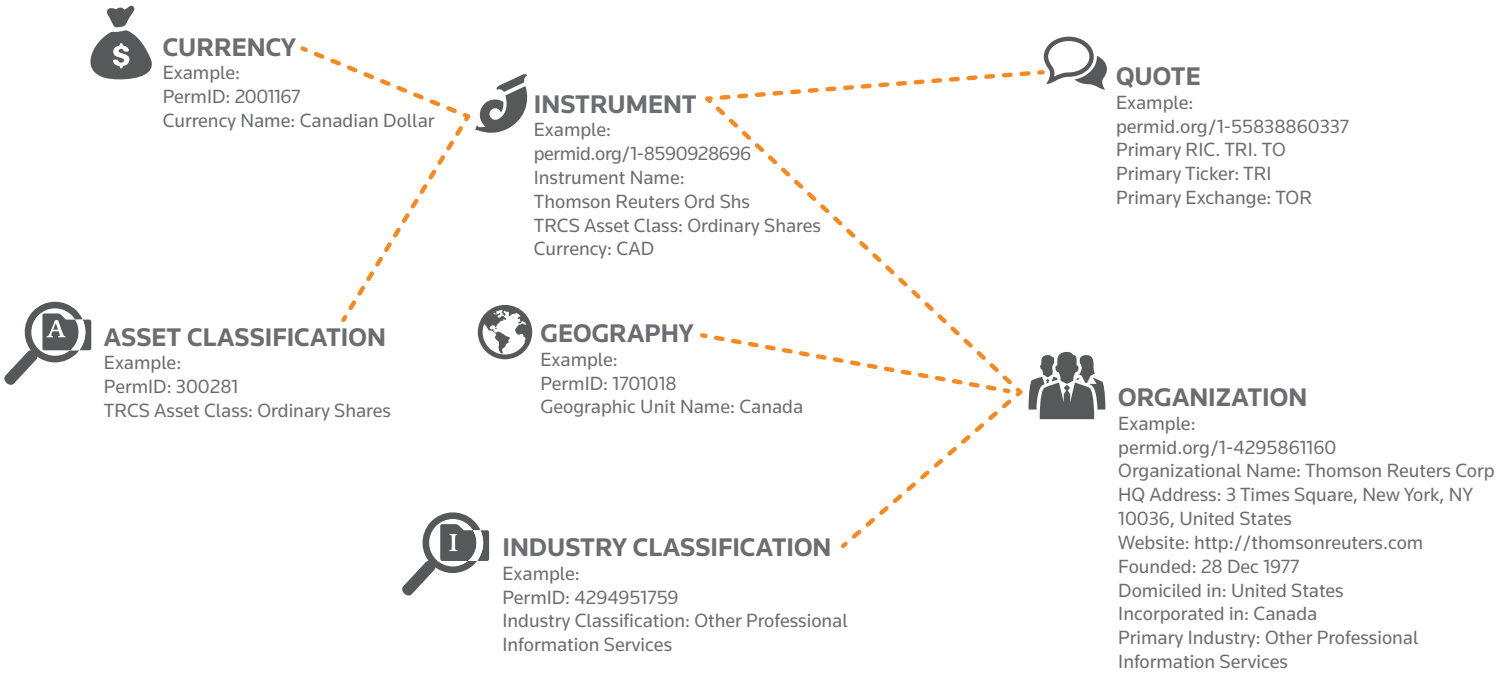
THE INFORMATION MODEL

The Information Model is a representation of the real-world things or concepts that Thomson Reuters can uniquely identify. It is a framework for describing content so that it can be delivered and reused in a variety of ways. It's important to note that the Information Model is not a type of data model.

The diagram below shows a small sample of the different types of objects identified by Thomson Reuters and the defining characteristics captured for each instance.

Not every object identified is published as open data today, therefore the diagram highlights where an Open PermID has been published vs. where a PermID has been assigned and is currently being used internally only.

Information Model Example



The Information Model captures the distinct sets of objects for which there is an authority, whose members' defining characteristics are captured and uniquely identified.

It also captures dependency between the defining characteristics of objects maintained by different authorities (e.g., instruments and organizations).

All objects managed, will be of an agreed "type":

- An entity is anything that needs to be uniquely referenced, in order to refer to it, reason about it, describe some aspect of it or use it to describe something else.
- A relationship is an association between two information objects and acts as a connection or navigation path between them.

The different types of object are agreed within the community using exactly the same criteria, for example, for an entity:

- The object must be externally recognized by participants.
- The object should have at least one associated public identification scheme.
- The object should have a legal status.

Other object types are supported internally and may well be exposed externally in the future.

HOW IT WORKS

There are two major elements to the operation of PermIDs: population of and reference to the federated registry.

THE FEDERATED REGISTRY: THE AUTHORITY AND THE MASTER DATABASE

An authority is an organization or group responsible for maintaining defining characteristics about – and assigning PermIDs to – some set of objects in the real world within a registry on behalf of the entire community.

Rather than one instance of authority and registry for all types of objects, we federate the operation, typically with distinct authorities managing registries for distinct types of objects. For example, the 'Organization Authority' manages a registry of organizations.

Each registry has a single point of physical storage and resolution from which all use of PermIDs is synchronized. This is called the master database.

(Note that the authority is responsible for capturing the existence of the right set of objects described, not for curating any other information properties or characteristics that describe those objects. Properties and characteristics may be subsequently associated with the object, added by other groups and systems using the PermID as their reference point.)

Population of a Registry

The goal of the authority is to ensure that the registry contains entries that reflect

- The state of the objects of description in the real world.
- The information interests of the community of PermID users.

This requires that the authority understand the community and the available sources of information and also seek continuous, explicit input from those sources and from the community as to new needs, changes in the real world, data corrections that may be required, etc.

The authority 'curates' a set of identities, each one affirming the existence of a corresponding object of description, either in the present time or in the past. For each identity, the authority must maintain sufficient breadth of 'defining characteristics' such that anyone in the community can confidently reconcile their identification of an object with the corresponding PermID. The authority will:

- Use a variety of information sources, open and closed, Thomson Reuters and third-party.
- Include in the set of maintained 'defining characteristics' a range of symbols, names and other identities by which the object might also be known within the community.
- Use their specialist knowledge to reconcile distinct objects (one reason for federating authorities based on type of object).
- Assign each discrete new object a PermID (from a ranged set rather than algorithmically, to avoid possibility of duplicates).
- Use understanding of the community, seeking continuous, explicit input from the community as to new needs, changes in the real world, data corrections that may be required, etc.
- Use a standard administrative life cycle (Appendix B) to maintain the administrative state of each object (e.g., live, obsolete, superseded, etc.) across all types of objects.
- Detect and resolve any duplicate or erroneous objects (using the administrative life cycle: once created, objects and their PermIDs are never deleted).
- Ensure that changes in the state of the registry are timely as required by the community of users.

Life Cycle of an Object

Since information changes over time, the authority maintains both administrative and native life cycle states for each object.

- The administrative life cycle describes the relevance and currentness of an object; it captures a full history of the changes in state in the lifespan, starting with its creation. The administrative life cycle is common for all objects.
- The native life cycle is independent of the administrative life cycle and depicts the evolution of an object over a period of time. The native life cycle is unique to each object type.

Information Rights

Since the authority uses some licensed, third-party sources as a basis to establish and maintain some registry entries, changes in owner license may require Thomson Reuters to remove any entries based fully on their data. If this happens, we will use an administrative life-cycle state to reflect this and 'logically delete' the entry – 'nulling' all of the values of the defining characteristics. The registry entry and the PermID will continue to exist as a 'null' reference point to avoid dangling references and to allow any such references to retrieve the status of the entry.

The authority will work to re-establish the entry based on other sources. If this is possible, the authority will attempt to ensure the same PermID is used. If it cannot, it will be superseded in the standard manner.

IDENTITY REFERENCE SERVICES

A small set of standard services are made available to the user community, each operating across the full universe of described objects (i.e., the aggregate set of objects published by the federation of registries):

- **Search:** Essential for finding the right PermID to use by passing in values for some subset of the 'defining characteristics'. One or more registry entries each with their respective PermID may be returned: by examination of the full set of defining characteristics for the entries, the user may select the appropriate PermID. Search can operate 'by type' or across types, by using common attributes such as name.
- **Reconciliation:** Cross-referencing between one or more PermIDs and corresponding identifier or symbol schemes commonly used in the community. This will require exchange of some defining characteristics as well as the identifiers or symbols.
- **Dereferencing:** Lookup of the defining characteristics, status, metadata and other limited data elements including common relationships that all correspond to a specific PermID.
- **Feedback:** A means by which the community is encouraged and motivated to provide information on needs and data accuracy.
- **Bulk Feed:** A mechanism for obtaining large amounts of information, which in turn will enable caching closer to a third party.

Additional value-add services will be added in the future.

For the scheme to work accurately, users of identity services should:

1. Synchronise any copy of a registry entry with the Thomson Reuters master registry in a timely fashion.
2. Follow changes in the status of registry entries as communicated by the identity reference services. If an entry is superseded, references and relationships should be updated replacing the superseded entry's PermID with the superseding entry's PermID.
3. Search for registry entries using the common service or a synchronized privately cached copy.

Search, Reconciliation and Selection from a Candidate List

It is important to understand the nature and role of the search and reconciliation services. While dereferencing is completely deterministic (a PermID refers to a specific and unique object of description), search & reconciliation take a user-selected subset of 'defining characteristics' and identifiers or symbols and return one or more candidate entries in the registry (along with their PermIDs). The resulting user action to select the right candidate (or to note that there was no candidate that appears correct) is what decouples the user activities from those of the registry.

This compulsory activity on the part of the user might be performed by a human or more probably by some kind of machine intelligence – simple rules or AI algorithm of some kind.

Open PermID Interface

The Open PermID capability is provided on the Web free of charge and under a Creative Commons-BY open license (although some data, features and levels of access require user registration).

The Open PermID capability encodes and exchanges PermIDs as URIs with the following syntax: **permid.org/xxx-xxxxxxxxxx**

These URIs are intended to be completely opaque and persistent, containing no useful information in themselves.

The Open PermID capability comprises two elements:

1. A Web-based user interface intended for use by developers and data maintenance staff. It provides a search interface and value-added services for bulk, file-based reconciliation and object extraction from documents.
2. An interface allowing machine dereference of a URI-represented Open PermID into an RDF representation of the defining characteristics, state and metadata of the registry entry that corresponds to that Open PermID.

It should be noted that not all defining characteristics or all object types are supported through the Open PermID interface; the license provides more details.

Thomson Reuters Products

Support for PermIDs in existing F&R products will be added and extended over time.

Initially, Thomson Reuters desktop and feed products will simply add PermIDs to existing product models as a foreign key:

- Support will be limited to selected product platforms.
- PermIDs are encoded and supplied as 64-bit numbers (not URIs).
- Support is limited to use as a 'foreign key': PermIDs will be added as an extra column or field in order that they can be externally used to correlate objects across products.
- PermIDs are not supported as a retrieval key. Search UI interfaces may support lookup by PermID.
- None of the standard identity services (above) are explicitly provided through existing product interfaces.
- The full universe of objects of description and their corresponding PermIDs that exist in the federated registries may not be supported by existing products, which will continue to maintain their own distinct product models and coverage. Some object types may not be supported at all in some products. Coverage across product types and customer entitlement configurations will vary.

As a result we expect customers will use the Open PermID capability alongside existing Thomson Reuters products. While it is possible to convert a 64-bit PermID number to the URI syntax and vice versa, the difference between product and Open PermID coverage may mean that not all such conversions will be supported.

(A PermID obtained from a paid product converted to a URI format and dereferenced from the Open PermID capability may result in failed lookup [http 404]. A PermID URI obtained from the Open PermID capability converted into a 64 bit number may not be found using a specific product search interface.)

Subsequently, we anticipate

- PermID support across all other products
- Generic support for other types of objects (entities, relationships and events)
- Independent support for all standard identity services across full object universes, subject to access permissioning (independently of product model coverage)

We do not currently anticipate using the Open PermID as a content retrieval key. We do plan to use them as common means of reference, eventually helping us retire many of the proprietary symbologies we currently manage.

Access Permissioning

Some objects and attributes may be subject to third-party rights management. Where this is the case:

1. Access via Thomson Reuters products will be through permission in the usual fashion.
2. Access via the Open PermID interface will not be possible (failure to dereference an Open PermID because it has not been assigned to any object will be distinguished from failure to dereference due to access restrictions).

PRACTICAL USE OF THE OPEN PERMID FACILITY

Use in the Community

It is an explicit design goal that Open PermID will be used between participants in the community as a means of exchanging and communicating the identity of described objects. For this to work accurately, participants should

1. Communicate Open PermIDs in URI format.
2. Dereference Open PermIDs to check their status and meaning when passed and when received.
3. Dereference Open PermIDs using Thomson Reuters provided services or from cached copies that are maintained so as to be faithfully synchronized with the master versions.
4. Reconcile commonly used identifiers or symbols to Open PermIDs using Thomson Reuters reconciliation service.
5. When reconciling local data with Open PermIDs, ensure that
   a. The widest set of defining characteristics is used (to ensure best candidate set).
   b. The intelligence used to make selection from the candidate set is suitably expert.
6. Provide as much feedback as possible.

Extending the Community

In making the Open PermID facility open to all, an explicit goal is to dramatically increase the extent to which the objects whose identity is captured and the basis of that capture (the range of 'defining characteristics') is based on community feedback (rather than solely on internal Thomson Reuters decision-making processes). This is in addition to working for services and data to be provided under open license and according to the agreed best practices of the open-data community.

PermID minting is the process and mechanism currently being explored by Thomson Reuters to enable clients and partners to participate in the creation and maintenance of identities, therefore enhancing the ever-expanding universe within the community.

In due course, as we open up the process, we expect others in the community to take on the roles of the:

- **Contributor,** providing instance-level data on specific subsets of objects on a regular basis, thus enlarging the set of data from which Thomson Reuters manages the registry, and/or
- **Authority:** Managing specific instances of object (all defining characteristics, status and metadata) through a managed service provided by Thomson Reuters. In effect this equates to federating the authority that manages a set of objects. This could extend to one owner managing a whole set through Thomson Reuters provided services, such that the master system resides in Thomson Reuters while the authority resides in another organization.

Being an authority means managing identity for the whole community of users and sourcing and maintaining a corresponding range of defining characteristics, so managing the notion of existence and identity for everyone, not just locally. This will need acceptance from others in the community and may require extensions of trust that may not exist today.

We anticipate that the only unique role that Thomson Reuters retains longer term, in the extended community, is one relating to coordination; governance of standards and practices, maintenance of the Information Model, search and lookup facilities.

APPENDIX A TECHNICAL DETAILS

DESIGN GOALS

The Open PermID facility originally arose from an internal need in Thomson Reuters to work in a truly scalable, federated means yet have the result of our efforts – information – be usable as a set by our customers, in ways we understood well but also in ways we could not predict.

Some key goals of the facility include:

1. Maintaining the same form and method of identity irrespective of the subject or form of the information.
2. Ensuring that the precision of identity improves as the use of the facility increases, as a result of wider input about needs, relevance and quality.
3. The ability to consistently and accurately represent relationships and references between objects and information maintained by isolated groups.
4. The immutable identity of objects – once a PermID is assigned, the PermID cannot change. If the object changes, a new PermID is created and chain of supersedence is constructed to enable users to follow how it has changed. Objects and their PermIDs cannot be deleted, only marked as obsolete. Use of PermIDs for communication of historical information is therefore supported.
5. Bi-temporality; the model supports representation of valid time and transaction time where applicable.
6. Convergent consistency. The method is designed for federated environments in which general transactionality across sources is not relevant and consistency of information converges rather than being enforced. Consistency of identity (consistent state of the 'defining characteristics' for a given instance of object with a given PermID) is guaranteed.
7. The ability to be realistic and honest about change. The real world changes, and the immutability of objects and their related PermIDs ensures that we can reflect and record that change. It is also true that people and machines make mistakes. Corrections in identity (e.g., removal of duplicates) must be reflected consistently and swiftly and communicated to all PermID users in a timely fashion. A standard administrative life cycle that applies to all object types helps ensure that this happens.
8. Standards rather than systems. Being explicit about the rules of the method ensure that any implementation can join it as an authority or as a user.

The realisation that the method held a lot of value and promise for uses outside Thomson Reuters arose from the recognition that the original Thomson Reuters need is very similar to needs in other enterprises and in particular needs in connecting Open Data in the Web. The Open Data Institute and Thomson Reuters have published a new white paper, explaining how to use identifiers to create extra value in open data; Open Data Institute and Thomson Reuters, 2014, Creating Value with Identifiers in an Open Data World, retrieved from **thomsonreuters.com/site/data-identifiers**

ANATOMY OF THE PERMANENT IDENTIFIER

The following properties of the Open PermID facility are intended to be consistent, reliable and durable to users of the facility:

- **Defining Characteristics:** The characteristics of an object considered sufficient to establish a unique reference between different parties within the community.
- **Syntax:** The Open PermID is created as a URI, with the following syntax: **permid.org/xxx-xxxxxxxxxx**

  There is a numeric element to the Open PermID, which is referred to internally as a PermID and conforms to the same rules as the Open PermID. This will be exposed in our existing product models as a 64-bit number.
- **Context & Granularity:** permanent identifiers are intended to support their user community in terms of the range of defining characteristics supported and the granularity of their formation. As the community grows this may mean that objects previously regarded as one may divide to become two or more. The standard administrative life cycle will ensure that any such change is communicated to users consistently. However, granularity and scope will not shrink.
- **Scope:** permanent identifiers are not dataset specific: they are intended to be usable in any dataset maintained by the user community.
- **Authority:** permanent identifiers organize a mixture of coordinating authority function and community contribution.
- **Discoverability:** permanent identifiers are discoverable through dereferencing services.
- **Stability:** permanent identifiers are never reused and the objects they identify are never deleted from the registry.
- **Timeliness & Synchronisation:** The Open PermID facility is intended to be timely in terms of real-world change or notifications of administrative change. Synchronisation is in general supported through both push and pull methods, although Open services may not support push.
- **Temporality:** The Open PermID facility is intended to be able to support both post-corrected and point-in-time historical methods.

The critical properties of any identity scheme are explored further in the white paper, Creating Value with Identifiers in an Open Data World, referenced above.

APPENDIX B FEDERATED MASTERING PRINCIPLES

This appendix summarises the principles to which our internal authority systems MUST comply and which govern the behaviour and success of the Open PermID facility. They are published here, using Thomson Reuters internal terminology, as part of our initiative toward openness and transparency.

PRINCIPLE 1: INFORMATION OBJECTS

An information object is Thomson Reuters perception or 'surrogate' of the real-world thing or concept that it represents.

All mastered information objects will be of an agreed "type":

- An entity is anything we need to uniquely reference, in order to refer to it, reason about it, describe some aspect of it or use it to describe something else.
- A relationship is the association between two information objects and acts as a connection or navigation path between them.

Information object types are subject to approval.

Information object types will be recorded in a single registry.

Each information object type will be allocated a permanent identifier.

PRINCIPLE 2: INFORMATION OBJECT MASTER

A master is a database system that specialises in the management of a registry of the existence of an information object type and is the sole storage and maintenance point for those objects.

All instances of a given information object type will be managed by only one master.

The attributes recorded for a given information object can be mastered in multiple databases, but they do not all have to be recorded in the master.

PRINCIPLE 3: INFORMATION OBJECT AUTHORITY

Information objects are created, maintained, and owned by authorities.

An authority is an organisational unit, e.g., a group of content specialists, not systems, which create and maintain information objects. The authority is responsible for creating information objects within its own domain. The authority owns the content in the master and specifies the requirements for the system.

There will normally be one authority for a given type, but it can be federated and if so, all authorities for the type will collectively adhere to the principles.

An authority will not create an instance of an information object type that it does not own. If a master needs to reference an information object which does not exist and is of a type of which it is not the owner, it will request the owning authority to create it.

An authority will have a process to address the request for information-object creation from other systems and may build the functionality into the master.

PRINCIPLE 4: UNIQUENESS OF AN INSTANCE

An information object will not be created or registered without a set of attribute values that can be used to uniquely differentiate its existence.

Where there are multiple sources or there is a reliance on third-party content for an information-object type, there will be the ability to create a concordance and to highlight and reconcile any inconsistencies. This will prevent the creation of duplicate instances by an authority.

PRINCIPLE 5: PERMANENT IDENTIFIERS

When an information object is registered, it will be allocated a unique and unchanging identifier, a permanent identifier.

This is used to uniquely identify information object type instances across the entire namespace of Thomson Reuters.

Each information object created by a master is assigned a PermID from an allocated range which remains static throughout the life of the instance and can never be reassigned. PermID ranges will be allocated to mastering systems and not per information object type being mastered.

When ownership of an information object is transferred from one authority to another, the PermID is retained if there is no significant change to the object.

PRINCIPLE 6: PERSISTENCE

Once published, an information object will never be physically deleted and PermIDs will not be reused.

If an information object needs to be deleted, the record will be retained and the administrative states used to imply the logical deleted.

PRINCIPLE 7: ADMINISTRATIVE LIFE CYCLE

An administrative life cycle will be implemented for all information objects. The administrative life cycle focuses on streamlining the process for the capture and publication of the information objects within the masters, i.e., how they are handled in Thomson Reuters systems.

Data collection and maintenance actions will be segregated from the natural evolution of the entities.

An information object can be:

- **Created.** At this point a PermID is not allocated.
- **Registered.** A PermID is allocated.
- **Published.** All the defining characteristics are confirmed, a PermID has been allocated, and the content can be sent out in a strategic data interface.
- **Deleted.** This state can only be applied if the content has not been published.
- **Superseded.** Replaced by another one.
- **Suspended.** The master ceases to provide updates for the content, e.g., the supplier has withdrawn it or the information object is perceived to be of a different "type." For example , changing "organization" to "building," but no home for the new object type exists. This state is expected to be temporary.
- **Obsolete.** A logical deletion.

Masters will maintain and publish a full history of administrative life-cycle states for all information objects.

PRINCIPLE 8: NATIVE LIFE CYCLE

An authority should develop and implement a native life cycle for the information object types it masters.

The native life cycle of an information object is its internal life cycle based on its inherent nature. The native life cycle is independent of the administrative life cycle.

PRINCIPLE 9: TEMPORAL DATA SUPPORT

Mastering systems will support temporal dimensions for their content.

Authorities will determine the temporal dimensions required and which data items are managed on which dimensions, i.e., what we want to keep for business history and what we want to keep for correction history.

PRINCIPLE 10: ENTITY DE-DUPLICATION

Duplicate instances of the same entity will not be created and maintained. However, if one is created in error, it will be de-duplicated.

Masters will be able to remove the duplicate entity occurrences by selecting a survivor, merging data into the survivor, marking the other as superseded and recording the supersession relationship to the survivor. This means that the downstream system can always trace a superseded record to the one which replaced it.

PRINCIPLE 11: WITHDRAWAL OF INFORMATION OBJECTS

Masters can support the withdrawal of published objects in a non-disruptive manner. They will also be able to withdraw attributes.

Legal, privacy, and/or licensing obligations require that we are able to withdraw published content while ensuring that the consumer experience is not disrupted. Legal obligations related to retaining and distributing personally identifiable data vary across geographies. To support the withdrawal of such data after publication, we will ensure it is suitably identified as such.

The master will evaluate content to enable the identification and publication of any that is subject to possible withdrawal.

PRINCIPLE 12: RELATIONSHIPS

Relationships describe the manner of association between two independent, meaningful information objects. Each relationship will have a subject, an object and a predicate, each referring to an independent and meaningful information object.

Relationships connect information objects and form the basis of navigation between them. They will be defined unambiguously (in direction) and will be consumable by machines as well as interpretable by humans.

The existence of objects at the endpoints of the relationship will not be dependent upon the existence of the relationship. Alternatively speaking, an object that the relationship connects will have its own meaning and will be interpretable on its own in the absence of any other objects, including other relationships. Therefore, a relationship instance cannot be used to determine the uniqueness of information objects.

If a relationship should exist but complete information is not available, it may be published but will be marked as undetermined. In some cases there is a need to indicate the existence of a relationship to an entity where the entity is not known, but where there is semantic information that needs to be conveyed.

All "Published" relationship instances will have "Published" information objects at both its ends.

Some relationship types are mastered and behave as objects within the information model. Instances of objectified relationships are published with relationship permanent identifiers.

ABOUT THOMSON REUTERS

Thomson Reuters is the world's leading source of intelligent information for businesses and professionals. We combine industry expertise with innovative technology to deliver critical information to leading decision makers in the financial and risk, legal, tax and accounting, intellectual property and science and media markets, powered by the world's most trusted news organization. With headquarters in New York and major operations in London and Eagan, Minnesota, Thomson Reuters employs approximately 60,000 people and operates in over 100 countries. Thomson Reuters shares are listed on the Toronto and New York Stock Exchanges.

ABOUT THE AUTHORS

Bob Bailey
VP, Chief Information Architect,
Thomson Reuters

Bob is Chief Information Architect for Thomson Reuters. He has worked in the information industry for 30 years, for much of that time specialising in information for financial markets and investment management. He leads the Information Strategy team, who help focus cross-company efforts on new business approaches to information use, target information architecture and core information policy formation. Most recently this included leading an exploration of the potential for use of Linked Data and Open Data in and by Thomson Reuters, culminating in March 2014 with the company joining the Open Data Institute as a Partner.

Toni McDerment
Information Architect,
Thomson Reuters

Toni is an Enterprise Information Architect at Thomson Reuters. With 20 years experience in data management disciplines she has built up an invaluable knowledge of the News & Media, Investment Management, Securities Administration, Life & Pensions and Retail & Commercial Banking industries. Toni has worked at Thomson Reuters for nine years and is responsible for defining both the strategic and implementation aspects of the information architecture, orchestrating its execution and governing against it.

For more information

Send us a sales enquiry at
thomsonreuters.com/about/contact_us

Read more about our products at
thomsonreuters.com/products_services

Find out how to contact your local office
thomsonreuters.com/about/locations

THOMSON REUTERS™