ORCID Assertion Assurance Pathways[[1]](#footnote-0)

**Document History**

|  |  |  |
| --- | --- | --- |
| **Date** | **Activity** | **Decision** |
| 18 Jan 2018 | Document draft created |  |
| 10 April 2018 | Reviewed by Trust Working Group | Approved with minor revisions. |
|  | Revisions completed |  |
| 13 June 2018 | Document locked into comment only mode | Share with community (announced in [blog post)](https://orcid.org/blog/2018/06/13/assertion-assurance-pathways-what-are-they-and-why-do-they-matter) |
| 1Q 2019 | Next review cycle |  |

**Table of Contents**

[**Summary**](#_4k9kpxo1jthx) **2**

[**2. Definitions**](#_jdwi9ucswb8) **2**

[2.1 Assertions](#_j5dn46zp2wo) 2

[2.2 Persistent Identifiers](#_8okwjsh4l4t2) 3

[2.3 Items](#_dul2j3smv27i) 5

[2.4 Agents](#_5jm89q5iekaj) 6

[**3. Assertion composition**](#_udmr2cjsc0jv) **7**

[3.1 Example assertions](#_rclyonvhl602) 7

[**4. Assurance pathways**](#_n7tkcf9wnf55) **9**

[4.1 What does the assertion contain?](#_n05couvgzr5q) 9

[4.2 Who made the assertion?](#_c7kh5a899w17) 9

[4.3 Assurance matrix](#_t191f7lsotzq) 10

[**5. Recommendation**](#_6i5oqnmuvzym) **11**

[**Appendix**](#_29pp6imbh2mt) **11**

[Example identifier categorisations](#_f33vlanztbrj) 11

#

# Summary

The ORCID Registry enables connections between individuals (via their ORCID iD) and their activities and affiliations (via other identifiers and APIs). These connections are asserted by individual record-holders themselves, or by organizations they interact or otherwise are affiliated with. When an ORCID record shows 'Source: XYZ' we need to know whether XYZ is the individual, a service stating information on behalf of an individual or third party, or is information that XYZ created or is otherwise responsible for. Understanding how to properly evaluate assertions on the record is key to maintaining and building upon trust extended to ORCID by the research community.

This document defines the types of agents involved in making assertions and how the use of persistent identifiers (PIDs) can affect the nature of assertions. It then examines common assertion scenarios and how assertions can be categorised based on the agents involved and the PID used. Finally, the document examines and describes the assurance implications of an assertion sitting within a particular category.

# 2. Definitions

To facilitate a discussion of assertion assurance categories we need to have a common understanding of the terms used. This section describes identifiers that underpin assertions and the items identifiers reference, and defines the agents and relationships involved in making assertions.

## 2.1 Assertions

For ORCID, an assertion is a statement that connects a **Person**, with an **Item**, and includes information about the **Source**. Simple examples, currently supported by ORCID, include:

* **Assertion by Person:**  A Person enters their employment details directly into the ORCID user interface. The Person is listed as the Source.
* **Assertion by Member:** A Person submits an article to a journal, which collects an authenticated iD and update permission. When the article is published, the journal uses the permission to update the Person’s ORCID record, which lists the journal as the Source.

Conceptually, in addition to a **Person**, **Item**, and **Source**, an assertion should also include the **Assertion Origin** (the agent that made the connection between **Item** and **Person**), and also the **Item Origin** (the agent responsible for maintaining the **Item** metadata).

Merriam Webster defines assertions as:

the act of asserting or something that is asserted, such as:
a. insistent and positive affirming, maintaining, or defending (as of a right or attribute) an assertion of ownership/innocence.
b. a declaration that something is the case. “He presented no evidence to support his assertions.”

Based on this definition, all **Items** (see definition in Section 2.3) in the ORCID record are considered assertions.

Only record holders and ORCID members can write to the ORCID record. Therefore, record assertions can be attributed only to the record holder or a member. The ORCID iD can be used to make assertions outside of the record, which means we need processes to distinguish who made the actual assertion (Assertion Origin) with an ORCID iD and who put that assertion into the ORCID record (Source). “On behalf of” functionality is being developed to allow members to correctly reflect record assertion to either record holders or other ORCID members.See Section 2.4 for more on the people and organisations that make assertions.

Some items in the record do not (yet) have enough PID infrastructure to allow for the automatic calculation of assertion by the data consumer. In those cases, we may require additional metadata to make manual assurance easier. Record assertions with external PIDs are easier to do assurances on and over any given timeframe we should see PID infrastructure improve for all assertion items. See Section 2.2 for more on the identifiers in assertions.

## 2.2 Persistent Identifiers

At a basic level, a **persistent identifier (PID)** is exactly as you’d imagine - a reference to a ‘thing’ that can be used to uniquely identify it, in perpetuity. This is the basic requirement, but there are several other desirable PID characteristics which make some PIDs more useful for assertions than others.

The ODIN project defines a **trusted PID**[[2]](#footnote-1)[[3]](#footnote-2), which is unique, persistent, descriptive, interoperable and governed. The definition states that PIDs must:

* Be unique on a global scale, allowing large numbers of unique identifiers
* Resolve as HTTP URI’s with support for content negotiation, and these HTTP URI’s should be persistent[[4]](#footnote-3).
* Come with metadata that describe their most relevant properties, including a minimum set of common metadata elements. A search of metadata elements across all trusted identifiers of that service should be possible.
* Be interoperable with other identifiers through metadata elements that describe their relationship.
* Be issued and managed by an organization that focuses on that goal as its primary mission, has a sustainable business model and a critical mass of member organizations that have agreed to common procedures and policies, has a trusted governance structure, and is committed to using open technologies.

The ODIN definition of a trusted PID is quite similar to the definition of Findable, Accessible, Interoperable, and Re-usable (FAIR) data[[5]](#footnote-4). The FAIR principles relax the mechanism of interoperability, and so consider the technical requirement of having metadata available through http content negotiation to be overly prescriptive. They instead require that it be available from the same endpoint using a well-known process. Notably, they also expand the definition to include a requirement for provenance information, such as who created the metadata and when.

The FAIR Data Principles specify:

**To be findable:**

F1. (meta)data are assigned a globally unique and eternally persistent identifier.

F2. data are described with rich metadata.

F3. (meta)data are registered or indexed in a searchable resource.

F4. metadata specify the data identifier.

**To be accessible:**

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.

A1.1 the protocol is open, free, and universally implementable.

A1.2 the protocol allows for an authentication and authorization procedure, where necessary.

A2 metadata are accessible, even when the data are no longer available.

**To be interoperable:**

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles.

I3. (meta)data include qualified references to other (meta)data.

**To be reusable:**

R1. meta(data) have a plurality of accurate and relevant attributes.

R1.1. (meta)data are released with a clear and accessible data usage license.

R1.2. (meta)data are associated with their provenance.

R1.3. (meta)data meet domain-relevant community standards.

When the metadata attached to a specific type of PID meets the FAIR principles, we refer to it as a **FAIR PID**.

The concept of trusted and FAIR PIDs is something the community should be aiming for, but in many real world scenarios, one or more of these requirements is not met by commonly used PIDs. With this in mind, we also define a less onerous category for **resolvable PIDs**, a characteristic that makes them far more immediately useful than Basic PIDs. To be in this category we have required that the PID be resolvable universally and openly to some form of human readable data. That is, for a given value and PID type, a single set of rules can resolve every instance of that PID to a landing page, without requiring a subscription.

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Definition** | **Comment** | **Examples** |
| **Basic PID**  | A permanent and unchanging reference to an agent or item provided by a dedicated service. | These PIDs uniquely identify items, but cannot be universally resolved to the item or the item’s metadata. | URN, ISSN, ISBN, grant numbers. |
| **Resolvable PID** | PIDs that are either URLs, or can be transformed into URLs that resolve to human readable content using well known rules[[6]](#footnote-5).  | This category includes PIDs that are universally resolvable to the item, but without universally available item metadata.  | Handle, PURL, RFC |
| **FAIR PID** | PIDs that are resolvable, unique, governed *and* can be used to discover interoperable, well-defined metadata containing provenance information in a predictable manner. | These PIDs can be universally resolved to items and consistent item metadata using a well-known mechanism. | ORCID, DOI, Arxiv, Bibcode |

**Table 1. Defining PIDs**

## 2.3 Items

Items are the ‘things’ referenced by assertions, ideally by using one or more PIDs. Examples within the ORCID registry include works, peer review activities, funding, education, employment, alternative names, and websites.

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Definition** | **Comment** | **PID Examples** |
| **Item** | An activity or entity that a person created, used, or was affiliated with, such as works, peer review activities, funding, education, employment, alternative names, and websites. | Items may or may not be associated with a persistent identifier (PID).  | DOI, arXiv, handle, URL, GRID, LEI, etc.  |

**Table 2. Defining Item.**

## 2.4 Agents

People and organisations involved in making assertions are referred to as agents. These include the person the assertion is about, who made the assertion, who added it to the ORCID registry and who is responsible for the Item referenced by the assertion. Note that a single agent can play multiple roles.

Only record holders and ORCID members can write to the ORCID record. Therefore, record assertions can be attributed only to the record holder or a member. The ORCID iD can be used to make assertions outside of the record, which means we need processes to distinguish who made the actual assertion (**Assertion Origin**) with an ORCID iD and who put that assertion into the ORCID record (**Source**). “On behalf of” functionality[[7]](#footnote-6) is being developed to allow members to correctly reflect record assertion to either record holders or other ORCID members.

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Definition** | **Comment** | **PID Examples** |
| 1. **Person**
 | The individual who created or otherwise is directly involved with an item.  | In our model, persons are identified using ORCID iDs, and in the context of assertions the Person must also be the record owner. | ORCID iDs |
| 1. **Item Origin**
 | The person and/or organization that is commonly understood to be responsible for the source item and/or for managing the item’s metadata.  | This could be, for example, a journal or book publisher, an employer, a funder, a research facility, or a data center. In the case of self-publishing, this is the **Person**. | Organization ID or ORCID iD |
| 1. **Assertion Origin**
 | The person and/or organization that associated the item with the ORCID iD and is ultimately responsible for the contents of the assertion.  | This can be a **Person** or an ORCID member organization, and can be the same entity as the **Item Origin**. | Organization ID or ORCID iD |
| 1. **Source**
 | The person or organization that added the item to the ORCID record.  | This can be the **Assertion Origin** or can be a service provider acting “on behalf of”[[8]](#footnote-7) the **Assertion Origin**.  | Organization ID or ORCID iD |

**Table 3. Defining Agents.**

# 3. Assertion composition

An assertion should include fields for **Person**, **Item**, **Source**, **Assertion Origin**, and **Item Origin.**  (the agent responsible for maintaining the **Item** metadata) if either are different from the **Source** (the agent that added the assertion to the registry). At present, ORCID does not distinguish between **Source**, **Assertion Origin**, and **Item Origin.**. The consequence of this is that the information about how an assertion came into being is not being captured. For example:

* **Assertion by Person via an ORCiD member organization:** A **Person** adds information to a Member’s profile system, or chooses information to add to their record from an indexer. The system updates ORCID with itself as a **Source**, when in fact it is user-asserted data.
* **Assertion by Member via a Service Provider:** A member provides services to other members, and adds all their assertions about a **Person** with itself as the **Source**, when in fact the service provider is an intermediary.
* **Assertion by Member with derived data:** With the permission of a **Person**, an indexer adds information it has collected from other systems to ORCID, either derived algorithmically or manually chosen. The indexer is listed as the **Source** of the assertion, when in fact it has been derived from elsewhere.

Alongside information about the agents involved, assertions contain information about the referenced **Item**. This takes the form of minimal metadata about the Item within the registry, for example, if the **Item** is an article, the name of an article, and the **Item** PID(s), if present. PIDs serve a number of purposes; they enable discovery of the **Item**, they reference richer domain specific metadata and can contain provenance information about the agent responsible for the **Item** and its metadata (i.e., the **Item Origin** ). Without a PID, reliable discovery of **Item** metadata is made impossible for machines, and difficult for humans.

## 3.1 Example assertions

|  |  |
| --- | --- |
| **Description** | **Diagram** |
| Simple article submission:1. A **Person** authenticates to a Journal using ORCID and submits an **Item**.
2. The Journal hosts the **Item** **(Item Origin)**
3. The Journal creates the link between person and **Item** **(Assertion Origin)**
4. The Journal updates the ORCID registry **(Source)**
 |  |
| Searching an index for relevant works:1. A **Person** authenticates to an Index service using ORCID and searches for **Items**.
2. The Index finds **Items** in Journals **(Item Origin)**
3. The **Person** selects **Items** and asks the Index to link them with their ORCID record. **(Assertion Origin)**
4. The Indexupdates the ORCID registry **(Source)**
 |  |

**Table 4. Example assertions.**

These examples above highlight a gap in our understanding of the provenance of assertions within the ORCID registry. We are addressing these gaps with on-behalf-of functionality and the use of FAIR PIDs, as described in the subsequent sections of this document.

# 4. Assurance pathways

When considering the assurance underpinning an assertion, two of the fundamental questions about the composition of the assertion are required. “Who made the assertion?” and “What does the assertion contain?”. We will consider the implications of **Assertion Origin** and **Item Origin** as well as **Source** and **Item** here.

## 4.1 What does the assertion contain?

An assertion contains information about the connected item, possibly one or more PIDs and sometimes the nature of that connection.

The assurance of assertion content is influenced by the PIDs involved in that assertion. To identify People we rely on ORCID iDs, for organisations we use on of the various Organisation IDs coupled with their member ID, and for works we use any of the PIDs required by our members. We sort these PIDs into the three categories defined above; basic PIDs, resolvable PIDs and FAIR PIDs.

Basic PIDs enable humans to validate that the item described in a record is indeed the item referenced. Resolvable PIDs make this easier, require less specialist knowledge about the PID type and in some cases can be resolved to metadata automatically. FAIR PIDs make the process automatic, and enable the discovery of the Item Origin, that is, the agent responsible for creating and maintaining the item’s metadata.

Some items in the record do not (yet) have enough PID infrastructure to allow for the automatic calculation of assertion by the data consumer. In those cases, we may require additional metadata to make manual assurance easier. Record assertions with external PIDs are easier to do assurances on and over any given timeframe we should see PID infrastructure improve for all assertion items. See Section 2.2 for more on the identifiers in assertions.

In the absence of a FAIR PID, the Item Origin can sometimes be discovered through human investigation. Although not ideal, in practice this does help humans gauge categories of assurance. Comparing the journal title of an article within the registry with that on the publisher landing page is an example of this. Extra metadata such as dates and affiliations are often required to make this discovery process reliable.

If we do not know the Item Origin, it is difficult, or impossible, to understand the item itself. The presence of a PID, preferably a FAIR PID, makes the process easier.

## 4.2 Who made the assertion?

For assertions made by a **Person** about him/her/themself, directly to the registry, this is an easy thing to ascertain. However, with member assertions we move into more difficult territory. The scenarios above describe members adding assertions made by themselves, by a **Person** or by other members, yet there are currently very few ways of telling these apart. In many cases the **Assertion Origin** is the same as the **Source**, and indeed this has until now been the assumption made by the registry.

Without knowing the origin of an assertion, it is difficult to understand the assurance that can be ascribed to it. The **Assertion Origin** can sometimes be derived from FAIR PIDs automatically, where the PID metadata itself contains a pointer to the ORCID record. They can also be derived in a less robust and manual way, by looking for a reference to the **Person** in PID metadata or by comparing the **Source** to the information asserted. A good example of this is where the **Source** of an employment assertion in the registry is also listed as the employer.

## 4.3 Assurance matrix

Taking the various agents and PID types, we can build a matrix describing the characteristics of various combinations and how the assertion can be verified. The assurance level requirements do change with use context. However ORCID can make these general suggestions for additional assurances required for metadata in the record. We include assurance methods in *italics*, for the example where information from an ORCID record was used to populate an affiliation field in a form:

|  |  |  |  |
| --- | --- | --- | --- |
| Record assertion entity x PID types | **Person** | **Member** | **Member is item origin** |
| **No PID** | Manual assurance w/ Item Origin. *Affiliation institution should be contacted to verify details.* | Manual assurance w/ Person & Item Origin. *Affiliation institution should be contacted to verify details. Record owner should be asked to verify details are correct.* | Manual assurance w/ Person. *Record owner should be asked to verify details are correct.* |
| **Basic or resolvable PID** | Machine-assisted manual assurance w/ Item Origin. *PID metadata should make identifying affiliation easier. Affiliation institution should be contacted to verify details*. | Machine-assisted manual assurance w/ Person & Item Origin. *PID metadata should make identifying affiliation easier. Affiliation institution should be contacted to verify details. Record owner should be asked to verify details are correct.* | Machine-assisted manual assurance w/ Person. *PID metadata should make identifying affiliation easier. Affiliation institution should be contacted to verify details. Record owner should be asked to verify details are correct.*  |
| **FAIR PIDs w/ Authenticated ORCID iDs from certified member integration.** | Machine assurance w/ Item Origin. *PID resolution of metadata can be used to automate assurance.*  | Machine assurance w/ Item Origin. *PID resolution of metadata can be used to automate assurance.*  | *No additional assurance required.*  |

**Table 5. Assurance categories.**

An example of manual assurance with the Item Origin would be contacting a publisher (or researching their publications database) and having them confirm a particular person is associated with a particular article. Resolvable and FAIR PIDs make this process easier.

# 5. Recommendation

ORCID recommends we formalize the assurance pathway matrix, socialize with the research community, and ensure these assurance categories are transparent in the ORCID Registry UI and API.

ORCID will soon be separating out the agent adding something to the registry (**Source**) and the agent making the assertion (**Assertion Origin**). This will be through the introduction of on-behalf-of functionality, with which ORCID will start to record and expose this information in the registry. The link between **Assertion Origin**, ORCID members, and real world entities will be facilitated by Legal Entity Identifiers (LEI)[[9]](#footnote-8), which will become a required Organization ID for all ORCID members.

# Appendix

### Example identifier categorisations

**Handle**

While handles often resolve to pages with embedded Dublin Core metadata, its use is not mandatory and therefore not predictable. It is also impossible to search the entire handle namespace without resorting to a general purpose search engine. This means that handles do not come under the definition of trusted, rather they are considered Resolvable PIDs.

**URN**

Many URNs are resolvable, but not all. Of the subset that are resolvable, different rules apply depending on the URN namespace. These are considered Basic PIDs for our purposes.

**ISBNs**

It is impossible to resolve every ISBN through a single service. While WorldCat is an excellent and well governed source that does an excellent job of compiling a list of ISBNs and offers ways to transform ISBNs into metadata, it does not contain the complete ISBN catalogue. However, the vast majority that are resolvable are considered FAIR PIDs.

**Bibcode and Arxiv**

While not generally expressed as URLs, these are globally unique identifiers that can be resolved to metadata in a consistent manner from a single well-governed source. Bibcode supports HTML embedded metadata and Arxiv a form of content negotiation. They are considered FAIR PIDs our purposes.

**ORCID and DOI**

These meet all the criteria of a FAIR PID.

1. Previous version is available [here](https://docs.google.com/document/d/1TpUMjjlI8QdqjB0GW_gDnnc0_k6n4L1lUq_b8Gwl7Ck/edit#heading=h.88jjgjfj6rla). [↑](#footnote-ref-0)
2. ODIN Conceptual model of interoperability <http://dx.doi.org/10.6084/m9.figshare.824314> [↑](#footnote-ref-1)
3. ODIN Workflow for interoperability <http://dx.doi.org/10.6084/m9.figshare.1373669> [↑](#footnote-ref-2)
4. We have relaxed this requirement to: “must be possible to transform into URLs, be machine actionable and resolve to metadata in a well defined manner” [↑](#footnote-ref-3)
5. https://www.force11.org/group/fairgroup/fairprinciples [↑](#footnote-ref-4)
6. The rules used by ORCID for transforming identifier values into URLs can be found at <https://pub.orcid.org/v2.0/identifiers>. See also <https://identifiers.org>, which maps over 800 identifiers to URLs. [↑](#footnote-ref-5)
7. [On Behalf Of Feature Analysis White Paper.](https://docs.google.com/document/d/1qp2Sf_lbKk1bqrdBMCQalXBDELXQYjiJZ4eLoYoFjws/edit) [↑](#footnote-ref-6)
8. See On Behalf of [specification](https://docs.google.com/document/d/1qp2Sf_lbKk1bqrdBMCQalXBDELXQYjiJZ4eLoYoFjws/edit?usp=sharing) [internal] [↑](#footnote-ref-7)
9. LEIs are legal entity identifiers provided through GLEIF, a global non-profit entity that registers organizations through accredited local operating units. [↑](#footnote-ref-8)