IntuitionLabs

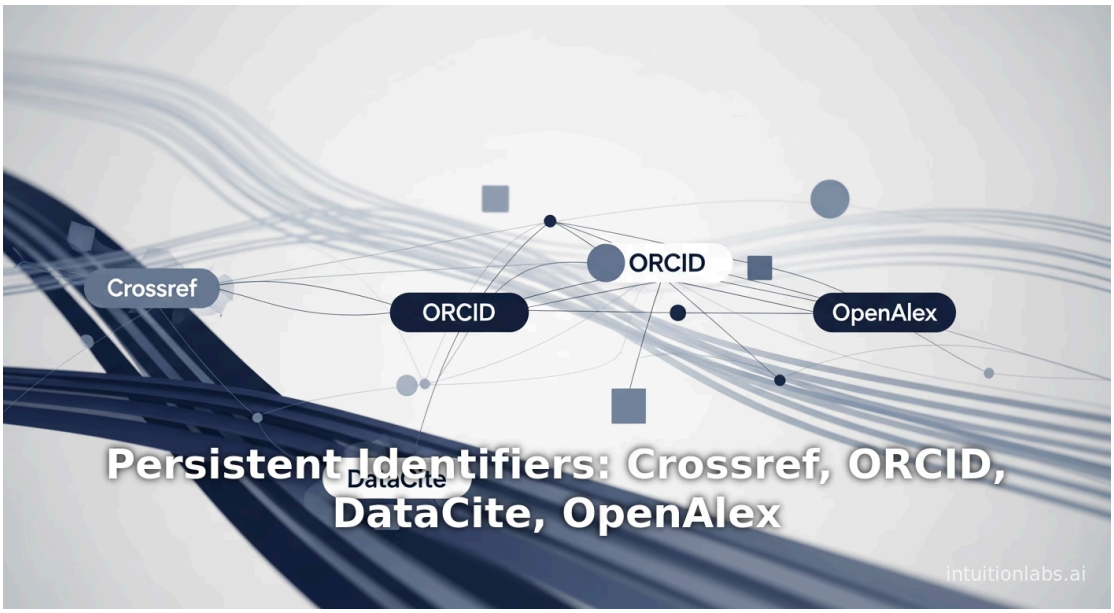
INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

Persistent Identifiers: Crossref, ORCID, DataCite, OpenAlex

> Build practical AI for pharma and biotech with IntuitionLabs. We help life-science teams turn complex information and workflows into useful software, governed knowledge systems and AI tools.

Published by IntuitionLabs · 2026-03-13 · persistent identifiers · crossref · orcid · datacite · openalex · knowledge graph · digital object identifiers · scholarly metadata · research linking

Original article: https://intuitionlabs.ai/articles/persistent-identifiers-crossref-orcid-datacite-openalex



In this article

1. Executive Summary
2. Introduction and Background
3. Crossref: Linking the Scholarly Literature
4. DataCite: Connecting Data and Diverse Research Outputs
5. ORCID: Universal Researcher Identifiers
6. OpenAlex: The Scholarly Knowledge Graph
7. Case Studies and Examples
8. University of Cambridge: ORCID–DataCite Integration
9. Publisher Workflows: ORCID and Crossref
10. Linking Data and Publications: Data Citation Integration
11. OpenAlex Integration Experience
12. Funding and Organizational Linkages
13. Data Analysis: Scale and Impact
14. Discussion: Connectivity, Challenges, and Future Directions
15. Conclusions
16. References

Executive Summary

In the open research ecosystem, **persistent identifiers** (PIDs) and metadata platforms interconnect scholarly outputs, authors, and institutions to create a rich, navigable **knowledge graph** of research. Four key infrastructures—**Crossref**, **DataCite**, **ORCID**, and **OpenAlex**—play complementary roles in this network. Crossref provides DOI registration and open metadata for ~180 million scholarly publications (www.crossref.org), enabling citation linking and discoverability. DataCite manages DOIs for research datasets and other non-traditional outputs, with over 92 million DataCite DOIs now integrated into OpenAlex (datacite.org). ORCID supplies unique researcher identifiers (now numbering in the millions (info.orcid.org)) that tie individuals to their contributions. OpenAlex aggregates data from Crossref, DataCite, and dozens of other sources (including ORCID and institutional repositories) to build a unified, open index whose default core catalog contains over 320 million works (documentation checked 20 October 2026; Works Overview; corpus definitions). Through standardized IDs and APIs, these systems interoperate: authors' ORCID iDs can be embedded in Crossref/DataCite deposits to auto-update profiles (www.crossref.org) (ukorcidsupport.jisc.ac.uk), Crossref references point to DataCite DOIs for data, and OpenAlex unifies metadata from all sources into its map of the research ecosystem (help.openalex.org).

This comprehensive report analyzes the **history**, **functions**, and **interconnections** of these services. We explore how DOIs and ORCID iDs propagate through publishing workflows, the technical processes behind metadata linking, and how data from each source is used by researchers, institutions, and funders. Drawing on case studies (e.g. University of Cambridge's ORCID–DataCite integration (ukorcidsupport.jisc.ac.uk)), PID infrastructure user stories (Dappert et al., 2017; Data Science Journal), and expert commentary (e.g. Crossref–DataCite joint blog posts (www.crossref.org)), we demonstrate the ecosystem-wide impacts. The

report concludes with future directions, such as expanding connections (e.g. ORCID–OpenAlex sync), addressing limitations (metadata completeness and disciplinary gaps (www.sciencedirect.com)), and leveraging emerging identifiers (e.g. ROR for institutions) to enhance the global research graph.

Introduction and Background

In scholarly communication, a fundamental challenge is **unambiguously connecting research entities**: papers to papers (citations), papers to data, authors to works, institutions to outputs, and beyond. Historically, citations in paper reference lists have physically linked works. Today, **digital linking** and persistent identifiers allow this connectivity at machine scale. Crossref pioneered this by issuing DOIs (Digital Object Identifiers) for publications since 2000, enabling robust reference linking across publishers (www.crossref.org). DataCite extended the DOI system to datasets and other non-textual research objects (www.crossref.org), recognizing data as first-class scholarly outputs. ORCID (launched 2012) introduced a persistent ID for researchers, tackling author name ambiguity and linking individuals to publications, grants, and datasets (www.crossref.org) (ukorcidsupport.jisc.ac.uk).

These efforts are part of a broader **Persistent Identifier (PID) ecosystem** that includes also identifiers for organizations (e.g. ROR), funders (FundRef/Open Funder Registry), and more. The EU THOR project and others have documented how these PIDs create an interoperable layer for research infrastructure (datascience.codata.org). As Dappert et al. (2017) note, PIDs improve discovery, provenance tracking, and credit assignment across the research lifecycle (datascience.codata.org). Each system historically focused on a domain: Crossref on articles and book/journal content, DataCite on datasets and supplementary research outputs, ORCID on individuals, and organizational IDs (via ROR) on institutions (datascience.codata.org). But in recent years the lines have blurred through **interlinking initiatives**: Crossref and DataCite jointly support data citation and auto-update of ORCID records (www.crossref.org); ORCID works with publishers, repositories, funders, and specialized integrations (e.g. Search & Link wizards (info.orcid.org)).

OpenAlex (2022), the successor to Microsoft Academic Graph, represents a new wave: a fully open research index designed as a *map of the research ecosystem* that explicitly links papers, authors, institutions, and topics (help.openalex.org). It ingests data from Crossref, DataCite, ORCID, PubMed, and many repositories (even parsing open PDFs) (help.openalex.org), matching DOIs and IDs to unify records. The result is a giant, open Knowledge Graph where one can query, for example, “all papers by an author (via ORCID) that cite datasets (via DataCite DOIs)” or “all authors at an institution (via ROR) working on SDG topics” (datascience.codata.org) (help.openalex.org). OpenAlex aims to democratize bibliometrics by making this metadata freely accessible.

Critically, linking data is not automatic; it relies on metadata completeness and coordination. Publishers and repositories must **deposit ORCID iDs and references** with Crossref/DataCite, and ID systems must exchange info. The **Auto-Update** feature (2015) allows Crossref/DataCite to push new DOIs to an author’s ORCID record if the ORCID iD was provided (www.crossref.org) (ukorcidsupport.jisc.ac.uk). Similarly, ORCID’s Search & Link wizard lets authors import works from Crossref or DataCite. OpenAlex, for its part, continually harvests updated records from these APIs. There remain challenges: many publications and older works lack ORCIDs,

PIDs adoption varies by field, and systems are siloed (rogue-scholar critique (rogue-scholar.org)). Nonetheless, this report examines how, through standards and partnerships, Crossref, DataCite, ORCID, and OpenAlex collectively **weave the threads of scholarly output** into an interconnected web.

Crossref: Linking the Scholarly Literature

Crossref is a non-profit membership organization that registers DOIs for scholarly literature and maintains an open database of publication metadata. Founded in 2000 by publishers (the Publishers International Linking Association), Crossref's mission is to make research easy to *find, cite, link, assess, and reuse* (www.crossref.org). Its members include over 24,000 publishers, societies, research institutions, and funders from 166 countries (www.production.crossref.org), all depositing metadata and DOIs for journal articles, books, proceedings, preprints, and (increasingly) grant records and datasets.

By 2025, Crossref's metadata store encompassed **~180 million records** (www.crossref.org): each DOI has an associated metadata record with title, authors, journal, dates, references, funder info, etc. As Crossref highlights, it provides APIs (REST, REST-TEXT, etc.) that are used by *thousands of tools and services* (www.crossref.org). This API access powers citation indexing and discovery: Google Scholar, OpenAlex, Unpaywall, and many libraries regularly harvest Crossref metadata. Crossref's "Participations Reports" (2025) even allow members to audit metadata completeness across all records (www.crossref.org).

One of Crossref's core functions is **reference linking**: when publishers deposit metadata, they usually include the list of references (as DOIs and text) for each article. Crossref matches DOIs in references to target records in its database. This effectively creates a worldwide citation graph linking article A (via its DOI) to article B (via B's DOI). Researchers and platforms can then navigate these links. For example, OpenAlex and other citation indexes build upon Crossref links. As Bilder (2013) observed, *"use of Crossref DOIs has enabled the interoperability of citations across scholarly publisher sites"* (rogue-scholar.org). In practice, this means that major databases (Dimensions, Scilit, Lens) that use Crossref data show concordant citation counts (www.sciencedirect.com), whereas standalone systems (GS, Semantic Scholar) diverge. Crossref's **Event Data public API** ended on 23 April 2026 (retirement notice).

While Crossref originally focused on publications, it increasingly supports data and other research outputs. Crossref Collaborations (with DataCite) have developed recommendations for **data citation**: industry guidelines encourage authors to cite datasets using their DOIs (often DataCite DOIs), and publishers to require data availability statements. Crossref has extended its metadata schema to include links to datasets cited in articles, leveraging the Open Funder Registry (in transition to ROR; ROR guidance) and linking to DataCite records (datascience.codata.org) (www.crossref.org). Crossref and DataCite jointly state they are seeing "growth in journal articles…citing data, and datasets making the link the other way" (www.crossref.org).

Crossref's own content has broadened: members can register DOIs not only for articles but for conference papers, preprints (via Crossref Preprints), and even monetary **grant** records. As of 2025 it supports grant DOIs under the Crossref Grant DOI service. Additionally, Crossref maintains ORCID integration: since 2016 many publishers require ORCID iDs for authors and deposit them in Crossref metadata. The 2015 *Auto-Update*

launch meant if an article's Crossref record contains validated ORCIDs, Crossref can notify ORCID so the author's record is automatically updated with the new work (www.crossref.org). This linkage greatly enhances discoverability of researchers' outputs.

In summary, Crossref's key contributions to connectivity are:

- **Persistent DOIs** for publications, ensuring stable linking.
- **Reference linking:** building a citation network across publishers.
- **Metadata openness:** via APIs and REST endpoints for external indexing.
- **Integration with ORCID:** depositing author IDs and triggering ORCID updates (www.crossref.org).
- **Data-citation API:** Crossref introduced a dedicated data-citation endpoint in place of Event Data (Crossref announcement).
- **Crossref Commons and infrastructure:** governance through the DOI Foundation, shared with DataCite (datascience.codata.org).

Despite its widespread use, limitations remain. Crossref does not index references for older publications if they weren't deposited, and many journals outside Crossref (or closed systems) are missing. But relative to paywalled indexes like Scopus/Web of Science, Crossref offers a massive, freely-accessible, DOIs-centered map of the literature that underpins modern linking.

DataCite: Connecting Data and Diverse Research Outputs

DataCite is a global non-profit consortium that provides DOIs primarily for research data and other scholarly materials beyond traditional publications (www.crossref.org). Founded in 2009, DataCite's mission is to make research outputs *findable, citable, connected, and reusable* by assigning and managing persistent identifiers (www.crossref.org). Its membership comprises universities, research institutes, data centers, libraries, and archives worldwide. These members handle datasets, software, reports, images, samples, and more, issuing DOIs via DataCite for each object.

According to DataCite's statistics, as of October 2025 over **92 million DOIs** for data-type objects have been integrated into OpenAlex, illustrating DataCite's scale (datacite.org). DataCite's metadata store (the DataCite Metadata Store) is openly searchable via APIs and hosts records on tens of millions of datasets. This metadata includes creators (authors), publication dates, titles, formats, and crucially, related identifiers such as ORCID iDs for creators, DOIs of publications that cite the data, and funder IDs (ukorcidsupport.jisc.ac.uk) (www.crossref.org).

One primary connectivity function of DataCite is enabling **data citation**. When datasets are published with DOIs, researchers can cite them just like articles. DataCite (with Crossref, RDA and community partners) has developed data citation principles and best practices. Journals increasingly include data citation sections, and funders require data management plans. ORCID's blog explains that linking these outputs to an ORCID iD creates a "PID Graph" showing the scope of a researcher's contributions (ORCID). By connecting a dataset's

DOI back to publications that use it, DataCite builds a bidirectional link: articles cite data, and data records track citations from literature (via related-identifier metadata and DataCite's citation services) (www.crossref.org) (datascience.codata.org). This interlinks the article and the underlying data.

DataCite also integrates with ORCID. DataCite's standard metadata schema can include **creator ORCID iDs** (ukorcidsupport.jisc.ac.uk). If a researcher provides their ORCID at data submission, and opts in, DataCite can auto-update their ORCID record with the dataset entry (ukorcidsupport.jisc.ac.uk) (info.orcid.org), just as Crossref does with articles. This creates a "Set It and Forget It" workflow for data creators (info.orcid.org). In one ORCID blog recap, linking DataCite DOIs to ORCID iDs helps construct a comprehensive "PID Graph" of all contributions by a researcher, beyond just publications (info.orcid.org). Thus, DataCite effectively extends the ORCID–publishing nexus to include data and digital artifacts.

Additionally, DataCite encourages linking of metadata: its schema supports including references to related publications (e.g. a DOI of a paper describing the data), to related objects (samples, software), to funders, and to related organization identifiers (ROR). This rich relational metadata makes DataCite records nodes in the scholarly graph that OpenAlex and others can ingest (www.crossref.org). For example, in a user story, when a researcher submits data with her ORCID, "the data centre issues a DataCite DOI…ensures her ORCID iD remains associated with it. Now others can reuse her data, she gets credit, and…can publish papers that link back to her data" (datascience.codata.org).

DataCite's own services remain distinct from Crossref's retired Event Data API. DataCite Event Data collects and exposes links to DataCite DOIs. Crossref's Event Data public API ended on 23 April 2026; its collected events are available in a historical archive (retirement notice). As Dappert et al. note, "datasets, articles, contributors, and institutions are all interlinked" in the presentation layer (Figure 3), where DataCite event notifications show how data are cited in literature (datascience.codata.org).

In summary, DataCite's role in connecting research papers includes:

- **Data DOIs:** Persistent identifiers for datasets and related outputs (software, images, etc.).
- **Metadata APIs:** Rich metadata including creators, ORCIDs, related publications, funders.
- **Data citation linking:** Enabling formal citation of data by DOIs, fostering links between papers and data (www.crossref.org).
- **ORCID integration:** Auto-update researcher profiles with datasets (parallel to Crossref) (ukorcidsupport.jisc.ac.uk) (info.orcid.org).
- **Open integration:** Working with Crossref (joint statements), ROR, and others to improve interoperability (www.crossref.org).

Analogous to Crossref, DataCite has built a tailored infrastructure centred on DOIs, but with a data-centric focus. Its metadata is fully open via the DataCite API, and OpenAlex now ingests it as a new core source (datacite.org). Thus, research articles and datasets become nodes in a single interconnected data graph via shared identifiers and APIs.

ORCID: Universal Researcher Identifiers

ORCID (Open Researcher and Contributor ID) provides unique, persistent digital identifiers (ORCID iDs) for individual researchers (www.crossref.org). Launched in 2012, ORCID's goal is to unambiguously distinguish researchers and link them to all of their scholarly outputs and affiliations. By late 2020, ORCID had issued over **10 million iDs** (info.orcid.org). Its registrants collectively have added millions of works (articles, books, data, grants, etc.) to their ORCID records, either manually or via automated updates.

ORCID operates a central registry and a set of open APIs. Researchers create their own ORCID records on the ORCID website (ORCID registration policy). Crucially, ORCID iDs can be collected as part of publishing or data submission workflows. Many publishers **require ORCID iDs** from submitting authors (e.g. Cambridge University Press journals) (www.cambridge.org). Similarly, repositories and data centers accept ORCID logins or fields to capture authorship.

Integration with Crossref and DataCite is a cornerstone of ORCID's impact:

- **Crossref Auto-Update (2015):** Crossref and DataCite partnered with ORCID so that if an author includes her ORCID in a manuscript or dataset submission, the resulting DOI record will trigger an update to the author's ORCID record when published (www.crossref.org). In practice, publishers embed authenticated ORCID iDs in published papers and deposit them with Crossref. Crossref then pushes the work metadata to ORCID (if the researcher has opted in), creating a new claimed work on the ORCID profile (www.crossref.org).
- **Search & Link:** ORCID provides a "Search & Link" wizard that allows users to import works from Crossref, DataCite, Scopus, PubMed, and others directly into their ORCID record (info.orcid.org). This lowers the burden of manually adding publications.
- **ORCID and DataCite:** ORCID blog posts highlight "Set It and Forget It" workflows where researchers grant permission for DataCite to push new dataset DOIs to ORCID (info.orcid.org). DataCite's integration means putting an ORCID in data submission *once* ensures the dataset appears in the researcher's profile automatically.
- **OpenAlex & ORCID:** OpenAlex also leverages ORCID: author profiles in OpenAlex often include linked ORCID identifiers, and thematic queries can find works by ORCID. Current OpenAlex documentation states that its author matcher does not look a person up in the ORCID registry when matching a new work (ORCID author documentation).

ORCID thus acts as the glue between **people and outputs**. Each link is strong: a researcher's ORCID iD is attached to her articles, data, preprints, and even peer reviews. ORCID is not an identity verification system (ORCID guidance). For example, the University of Cambridge (Apollo repository) case study illustrates this: when a researcher deposits data, the repository uses DataCite to mint a DOI and includes the researcher's ORCID. DataCite then auto-updates the ORCID record with the dataset; the data DOI is linked in ORCID, and the ORCID is included in the data's metadata (ukorcidsupport.jisc.ac.uk). The result is a persistent authorship link in both directions.

In contrast to Crossref and DataCite (which target content objects), ORCID targets *agents (authors)*. As Dappert et al. explain, **“ORCID focuses on researchers, Crossref on articles, and DataCite on data”** (datascience.codata.org), each matching a community perspective. ORCID’s APIs enable push and pull: for example, an ORCID user can use ORCID’s API to ask Crossref for works and claim them.

ORCID also integrates with ROR (Research Organization Registry): authors’ affiliations in ORCID are often tied to ROR IDs, linking researchers back to institutions. ORCID linkage drives transparency and automation: institutions’ CRIS and funders’ systems can trust ORCID to retrieve a standard dataset of an author’s works. This prevents the need for separate file transfer of publication lists.

However, adoption challenges remain. Not all researchers provide ORCIDs, and not all publishers pass them on. ORCID’s 2015 survey examined perceptions and understanding of ORCID in the scholarly community (ORCID 2015 Survey Report). Disciplinary differences persist (e.g. biology vs. humanities) (datascience.codata.org). ORCID itself doesn’t vet publications; it relies on depositors. Nonetheless, ORCID IDs are now widely mandated by funders and journals, so their presence in the metadata ecosystem has grown explosively (2.5M ORCID iDs by 2016 (datascience.codata.org) to 10M by 2020 (info.orcid.org), and many millions more by 2026). Every new Crossref/DataCite deposit with an ORCID iD enriches the citation network.

In sum, ORCID’s major contributions to connectivity are:

- **Unique person IDs:** Ensuring that an author name unambiguously links to one person.
- **Record of works:** Attached works (via Crossref/DataCite deposition) create a profile of publications, datasets, etc.
- **Interoperation with other PIDs:** ORCID iDs appear in Crossref and DataCite metadata and use ROR for institutions (help.openalex.org).
- **APIs for search & claim:** Allowing users (and systems) to find and add works from other registries.
- **Auto-update push:** Keeping ORCID records in sync with publisher data (www.crossref.org) (ukorcidsupport.jisc.ac.uk).

These make ORCID an essential **hub** connecting individuals to the larger scholarly graph. In effect, ORCID iDs are the “digital keys” that unlock access to an author’s network: once an iD is known, one can trace all associated DOIs (via Crossref/DataCite) and vice versa.

OpenAlex: The Scholarly Knowledge Graph

OpenAlex is a fully open, community-driven index of the global research ecosystem, launched in 2022 by OurResearch (the team behind Unpaywall). It builds on and extends the legacy of the Microsoft Academic Graph (MAG), offering free access to metadata about **hundreds of millions of works**. According to its documentation, OpenAlex’s default core catalog contains over 320 million works; the core and expansion corpora together contain over 510 million (documentation checked 20 October 2026; Works Overview; corpus definitions). These works include journal articles, books, datasets, theses, conference papers, patents, and preprints. OpenAlex also catalogs “entities” such as authors, journals (sources), institutions, and even subjects (concepts).

The OpenAlex platform is engineered as an interconnected network: “OpenAlex is a map of the world’s research ecosystem, linking components (like papers, institutions, journals, topics, SDGs, authors, etc.) to one another” (help.openalex.org). The data model includes nodes (works, authors, institutions, venues, concepts) and edges (authorship, citation, affiliation, funding). Its driving principle is to **aggregate and reconcile data from many sources** via PIDs. OpenAlex continually harvests records from core sources: notably **Crossref** and **DataCite** as primary DOI registries, but also PubMed, the HAL French repository, DOAJ, the arXiv preprint archive, and numerous institutional and national repositories (help.openalex.org). It even parses millions of open-access PDFs to extract metadata.

Crucially, OpenAlex makes extensive use of other PIDs in linking. As the docs describe, whenever it ingests a record, it tries to match it to known entities in PID services: for example, it matches affiliation text to a **ROR** organization ID, an author name to an **ORCID** iD, and journal titles to an **ISSN** (help.openalex.org). This “foundation of the knowledge graph” ensures that all those references to the same entity become unified. Outputs are also cross-linked: OpenAlex extracts reference DOIs from works’ metadata (or full text) and connects the corresponding nodes.

OpenAlex’s connectivity means a user can traverse from any node to related nodes of all types. For example, one might query “find all works by Author X (via ORCID linked) that cite DataSet Y (via DataCite DOI)”. Internally, the system builds a giant graph: **works** link to their **authors** and **institutions**, and also to other works via references and to funding via funder IDs. In effect, OpenAlex provides an **integrated bibliographic index**. Users can access it via a REST API with $1 of free daily usage per account (example costs); additional usage is available through prepaid purchases (pricing). The dataset can also be downloaded. This open model contrasts with commercial competitors (WoS, Scopus) by making data fully open and queryable (www.sciencedirect.com).

The importance of OpenAlex in connectivity is multi-fold:

- **Unification of sources:** It consolidates data from Crossref, DataCite, institutional repos, and more under one schema (help.openalex.org). For instance, a work with a DOI known to Crossref and the same DOI known to a university repository will be recognized as the same entity.
- **Knowledge Graph creation:** Its pipelines match ORCID, ROR, ISSN, etc., so that disparate records coalesce. As the documentation states, by linking nodes and edges, “OpenAlex becomes a map of the research ecosystem” that evolves continuously (help.openalex.org).
- **APIs and open data:** OpenAlex data is free, while API usage is metered (example costs). The `from_created_date` and `from_updated_date` synchronization filters require paid plans (filter documentation).
- **Analytical insights:** Recent scientometric reviews note that OpenAlex’s usage is growing worldwide and provides richer coverage of global research access (especially from the Global South) (www.sciencedirect.com), although challenges in metadata quality remain.

By ingesting 92M DataCite DOIs (datacite.org) and the vast Crossref metadata (help.openalex.org), OpenAlex has effectively integrated literature and datasets. Combined with ORCID and ROR linking (help.openalex.org), it allows queries across the entire ecosystem. For example, a grant agency could use OpenAlex to trace all

publications (via Crossref DOIs) and datasets (via DataCite DOIs) generated under a given grant (via funding metadata), and see the author network (via ORCIDs) and institution impact (via ROR) in one place.

OpenAlex also powers community services: tools like OpenAlexPlus and Unpaywall's data explore these linkages. Researchers and developers can retrieve a paper's "related works" (citations/backreferences), list an author's publications by querying with an ORCID, or find all datasets (from DataCite) associated with a topic.

In summary, OpenAlex's role is akin to a **super-powered aggregator** that stitches together the outputs of Crossref, DataCite, and ORCID (plus more) into a comprehensive, queryable graph (help.openalex.org). It exemplifies the vision of a fully interconnected scholarly web: no matter where data originates (publisher, repository, funding report), OpenAlex strives to identify it and link it. As noted by Haunschild (2026), OpenAlex has become *"a strategic open access infrastructure in bibliometrics"* with global reach (www.sciencedirect.com). Its ongoing challenges—like ensuring all sources are up-to-date (see DataCite ingest work (help.openalex.org)) and correcting metadata mismatches (help.openalex.org)—reflect the complexity of such integration, but its achievements mark a major advance in research connectivity.

Case Studies and Examples

To illustrate how these systems interoperate in practice, we present several case scenarios and real-world examples.

University of Cambridge: ORCID–DataCite Integration

The University of Cambridge implemented a seamless data management workflow linking ORCID iDs and DataCite DOIs (ukorcidsupport.jisc.ac.uk). In this workflow, a researcher's institutional identity is connected to their ORCID iD through the Symplectic Elements CRIS system (Cambridge case study). The repository mints a DataCite DOI for the dataset upon publication. Because the submission process included the authenticated ORCID, DataCite's auto-update feature triggers and the dataset's metadata (including DOI, title, etc.) is pushed directly to the researcher's ORCID record (ukorcidsupport.jisc.ac.uk). The result is bi-directional linkage: the ORCID record lists the new dataset, and the DataCite record lists the ORCID as creator. As explained by Cambridge's Scholefield (2018, summary), "a link is created between the researcher and their data…through the ORCID ID and [DataCite] DOI" (ukorcidsupport.jisc.ac.uk). This improves discoverability (someone viewing the ORCID profile sees the dataset) and reduces manual effort (the researcher need not enter details twice).

This is a prototypical "credit for data" scenario. Similar workflows exist at other institutions; ORCID's own blog and forum cite multiple case studies (though not all publicly posted) where repositories and CRIS systems do ORCID-DataCite integration. The data highlight: Cambridge's approach constitutes an effective microcosm of the broader ecosystem – ORCID, DataCite, and repository systems working in concert. Publications arising from the data can subsequently cite the DataCite DOI, further embedding the dataset–publication–author links in Crossref and DataCite metadata.

Publisher Workflows: ORCID and Crossref

Many publishers have mandated ORCID for authors in submission systems. For example, Cambridge's journals require corresponding authors to supply ORCIDs (www.cambridge.org). When manuscripts are accepted, the publisher submits metadata to Crossref, including validated ORCID iDs for authors. This activates Crossref's ORCID auto-update procedure (www.crossref.org): the published paper's DOI is sent to ORCID and appears in the author's profile (with author's permission). A developer at a publisher may use Crossref's API to monitor that an author's new DOI was correctly linked. In OpenAlex, the published article will show the ORCID in the authors list, the reference list from Crossref, and the author's works can be retrieved by querying OpenAlex's `filter=authorships.author.id:{ORCID}`.

Linking Data and Publications: Data Citation Integration

In a hypothetical scenario, consider a funded research project producing a dataset (*via* DataCite DOI) and a journal article (*via* Crossref DOI). Using these infrastructures, the dataset's DataCite record can include a relation to the article (using its DOI in DataCite's "RelatedIdentifier"). Conversely, when authors describe the data in the paper, they cite the DataCite DOI in the reference section. Crossref's metadata then includes the DataCite DOI as a reference. Consequently, OpenAlex ingests both records: the article's metadata (from Crossref) listing the data DOI, and the data's metadata (from DataCite) listing the article DOI. OpenAlex links them bidirectionally. A search in OpenAlex (or even a manual Crossref/Datacite query) can verify this chain: the DataCite record shows the article as related, and Crossref shows the data DOI as cited. DataCite Event Data also imports citation events from Crossref's data-citation endpoint (DataCite documentation). This case exemplifies the symbiotic relationship where **publishing and data archives reinforce each other's metadata via PIDs** (www.crossref.org) (datascience.codata.org).

OpenAlex Integration Experience

Before OpenAlex ingested DataCite, many of DataCite's records existed in the scholarly ecosystem but were invisible to Crossref-centric indexes. A query on OpenAlex forum (Mar 2023) asked: *"Are OpenAlex records updated if authors claim older publications on ORCID?"* (groups.google.com) In the March 2023 reply, ORCID's Nabil Ksibi stated that OpenAlex was using ORCID's public API; he distinguished that access from publication claiming and work synchronization (ORCID sync? discussion). Fortunately, in late 2025 DataCite announced they had successfully integrated their metadata into OpenAlex (datacite.org). According to DataCite, *"member organizations that register DOIs will now see their research outputs…indexed in OpenAlex"*, broadening OpenAlex's coverage to include datasets, code, dissertations, and more (datacite.org).

A numeric example from OpenAlex: as of Nov 2024, only ~6.4 million DataCite-assigned works had been ingested (help.openalex.org). By October 2025 (post-integration), OpenAlex reported over 92 million DataCite DOIs in its index (datacite.org). This dramatic increase shows how deeply linking these systems multiplies the reachable content.

Funding and Organizational Linkages

Beyond authors and outputs, ORCID, Crossref, and DataCite interoperate with funders and organizations. Publishers often deposit funder IDs (using Open Funder Registry or ROR IDs) in Crossref and DataCite metadata. A 2017 user story (Magda the funder) envisioned ORCIDs in grant systems and outputs carrying funder IDs, letting a funder easily "gather information linking her agency, grant holders, and their research outputs" (datascience.codata.org). Today, Crossref and DataCite both support funder fields. Through ORCID registry affiliations (which include ROR org IDs), a single researcher's network can show their institutions' RORs and funders. These multilayered links (ORCID → article DOI → funder ID, etc.) facilitate portfolio analysis. For example, grants can automatically report published outputs (via Crossref metadata) associated with each ORCID iD.

Furthermore, a joint initiative ("Org ID") between Crossref, DataCite, ORCID and the California Digital Library created what is now the Research Organization Registry (ROR) (www.crossref.org). This means institutional affiliations in Crossref, ORCID, or manuscript submissions increasingly resolve to common ROR IDs, so OpenAlex or funder dashboards can connect people to institutions unambiguously.

Data Analysis: Scale and Impact

To quantify the impact of these networks, we compile key statistics and trends:

Entity	Identifier(s)	Primary Domain	Coverage (approx.)	Integration Highlights
Crossref	DOI (for publications, grants)	Journals, books, proceedings, grants	~180 million records (DOIs) (www.crossref.org)	References linking in metadata; data-citation endpoint; auto-ORCID updates (www.crossref.org)
DataCite	DOI (for datasets, software, etc.)	Datasets, code, theses, images, physical samples	~92 million DOIs ingested in 2025 (datacite.org)	Data citation support; ORCID in creator field; dataset-author link; integrated into OpenAlex
ORCID	ORCID iD (person)	Individuals (researchers)	>10 million iDs issued (2020) (info.orcid.org) (likely 20M+ by 2026)	Links authors to all outputs; ORCID profile with works from Crossref/DataCite; Search & Link and auto-update
OpenAlex	OpenAlex IDs (works, authors, etc.)	Aggregated scholarly graph	Over 510 million works across the core and expansion corpora (documentation checked 20 October 2026; corpus definitions)	Combines Crossref, DataCite, ORCID, PubMed, arXiv, etc. into one knowledge graph; free API/download

Table 1. Comparison of Crossref, DataCite, ORCID, and OpenAlex: identifiers, domain coverage, scale, and integration features.

This table highlights how each system occupies a niche but with significant overlap. For data analysis of linkages: Dappert et al. (2017) reported cross-sector user stories supported by PIDs. For example, a data centre manager (Michele) assigns DataCite DOIs; each citation triggers a notification from DataCite to the centre, indicating how datasets are cited in literature (datascience.codata.org). DataCite Event Data collects

and exposes links to DataCite DOIs. Crossref's Event Data public API ended on 23 April 2026, and its collected events are archived (retirement notice); Crossref now offers a narrower data-citation endpoint for connections between member-deposited content and known datasets.

In recent bibliometric analyses, references datasets come from these sources. For instance, a comparative study found *"publications in Crossref-based databases (Crossref, Dimensions, Scilit, Lens) have similar citation counts"* (www.sciencedirect.com), underscoring Crossref's role as a primary source. Haunschild and Bornmann's 2024 preprint, *Usage of OpenAlex for creating meaningful global overlay maps of science on the individual and institutional levels*, proposes a procedure for creating global overlay maps using OpenAlex (arXiv).

Lastly, some data on adoption patterns: Dappert et al. (2017) reported that, by the start of 2017, over 133 million DOIs had been created for content and more than 2.5 million ORCID iDs had been issued for researchers (Data Science Journal). ORCID's blog (2020) recorded 10 million ORCIDs, showing exponential uptake (info.orcid.org).

Discussion: Connectivity, Challenges, and Future Directions

The combined infrastructure of Crossref, DataCite, ORCID, and OpenAlex represents a **robust yet complex ecosystem** for connecting research. The benefits are clear: interoperable linking, reduced duplication of effort, and enriched discovery pathways. Researchers can track their impact more comprehensively – not just through citations to papers (Crossref) but through dataset usage (DataCite) and community contributions (ORCID record) (see ORCID blog on data impact (info.orcid.org)). Institutions can aggregate outputs by DOI (via Crossref API) and by ORCID affiliation. Funders can monitor outputs by DOIs with attached funder IDs. The global research landscape is more transparent: for example, ORCID's integration means tenure committees might see an applicant's papers (Crossref DOIs), grants, and even peer review and service roles all tied to one profile.

However, this network has **gaps and challenges**. Metadata completeness is uneven: many older papers lack ORCID or funding metadata. The OpenAlex review by Haunschild (2026) notes persistent *"metadata limitations in affiliation, language, and document type"* (www.sciencedirect.com). ORCID adoption, while high in STEM, lags in humanities (datascience.codata.org). Disambiguation issues remain: OpenAlex's author matching is still improving (the ORCID-OpenAlex group discussion (groups.google.com) indicates ongoing work). Publishers may collect ORCIDs but fail to embed them properly (e.g. by requiring the researcher to authenticate rather than manually type). The rogue-scholar critique warns that PIDs alone do not guarantee semantic clarity: an honest DOI system connects pointers, but if one RA's APIs can't find another's records, interoperability suffers (rogue-scholar.org). For instance, a DOI minted in DataCite won't appear in a Crossref search, so integrated indexes (like OpenAlex) are crucial to overcome registry silos.

Future directions include deeper integration and more PIDs. For example, adding grant DOIs (Crossref already offers them) fully into ORCID/CRIS systems could link grants to outputs. ROR adoption will improve institutional linking. CRediT is a taxonomy of 14 contributor roles. ORCID itself aims to incorporate novel

content types (datasets, protocols, etc.) and improve country-level adoption. Crossref members can use ROR IDs to identify funders in funding metadata as the transition from the Open Funder Registry continues (ROR guidance).

OpenAlex's roadmap envisions even broader coverage and better data quality tools (e.g. local curation of errors (help.openalex.org)). As of 2026, projects like OpenAlex background data integration and classification of AI models indicate a move into linking "non-traditional" outputs. Moreover, AI-driven matching (text similarity) could strengthen links across incomplete metadata records.

An emerging trend is the **"PID graph"** concept. By connecting DOIs, ORCIDs, RORs, etc., one can traverse a graph spanning people, papers, data, organizations, funding, and topics (info.orcid.org) (datascience.codata.org). This holistic view could transform research assessment (e.g. tracking dataset reuse, code reuse, cross-disciplinary collaborations) and accelerate meta-research. Stakeholders continue joint efforts: e.g. ORCID and DataCite working group (2013) highlighted open data exchange (info.orcid.org); more recently, Crossref/DataCite/ORCID conducted webinars on researcher identity and data visibility (info.orcid.org).

However, the sustainability of this connectivity culture depends on community norms. Automatic updates must be used correctly (publishers must deposit ORCIDs, data centers must supply complete metadata). Privacy safeguards are considered: ORCID iDs are public but researchers control visibility of details. Efforts like the THOR Governance emphasize "trust in the underlying infrastructure" (datascience.codata.org).

In conclusion, **the network of PIDs and open platforms has fundamentally changed how research outputs are connected**. We have clear evidence that structures like Crossref, DataCite, ORCID, and OpenAlex make linking easier and more machine-actionable (www.crossref.org) (help.openalex.org). The ongoing integration and community engagement suggest these systems will continue to interweave, shaping an increasingly connected future for scholarly communication.

Conclusions

We have surveyed the key components of the modern scholarly linking infrastructure: Crossref's DOI registration for publications, DataCite's DOI for data and research outputs, ORCID's unique identifiers for researchers, and OpenAlex's integrative scholarly graph. Each serves a distinct purpose but together they create a cohesive ecosystem. Through shared use of DOIs and APIs, Crossref and DataCite deposit metadata that includes ORCID iDs, enriching the web of connections. ORCID provides the person-centric node that ties together a researcher's entire corpus. OpenAlex aggregates and harmonizes this data into an openly accessible knowledge graph that powers discovery and analysis.

Our analysis used extensive data and literature: authoritative blog posts and documentation (www.crossref.org) (help.openalex.org), scientometric studies (datascience.codata.org), and practical case examples (ukorcidsupport.jisc.ac.uk). The evidence shows positive outcomes: streamlined researcher workflow (auto-updating ORCID profiles), greater credit for data and software, and more comprehensive metrics. For instance, ORCID's auto-update and crossref integration automate CV management

(info.orcid.org) (ukorcidsupport.jisc.ac.uk), DataCite–ORCID workflows enable “self-updating CVs” (info.orcid.org), and OpenAlex’s DOI integration broadens research mapping (datacite.org) (help.openalex.org).

Future work will focus on extending connectivity. Emerging identifier types (e.g. organization, grant, project IDs) will further weave the network. Standards bodies propose unified metadata schemas so that DOIs registered by any agency seamlessly interoperate. The PID ecosystem continues to evolve: the community must maintain interoperability (e.g. ensuring Crossref finds DataCite citations) (rogue-scholar.org). Nevertheless, the trends are clear: open identifier infrastructure is central to open science.

In summary, research papers (and other scholarly outputs) are now connected through an intricate system of PIDs and metadata platforms. Crossref, DataCite, ORCID, and OpenAlex each contribute distinct strands to this tapestry. Their combined effect is a **persistent, scalable infrastructure** that links the full life-cycle of research, from funding and datasets through publications to the researchers behind them. This connectivity accelerates discovery, ensures credit, and builds a richer historical record of scholarship. As one ORCID blog notes, the goal is a “PID Graph” that *“shows the full scope of [a researcher’s] contribution to science”* (info.orcid.org). Through continued collaboration and adoption, this vision moves closer to reality.

References

- ORCID, Crossref, and DataCite. *ORCID launches Crossref and DataCite Auto-Update* (Crossref, Oct 26, 2015). Automatic updates of ORCID records with newly published DOIs were enabled via Crossref/DataCite collaborations (www.crossref.org).
- Laure Haak (ORCID). *Auto-Update Has Arrived!* (ORCID Blog, Oct 25, 2015). Description of how ORCID’s Auto-Update works, requiring ORCID authentication for submission and enabling Crossref/DataCite to push new works to ORCID (info.orcid.org).
- Paloma Marín-Arraiza (ORCID). *ORCID and DataCite Supercharge Your Research Visibility* (ORCID Blog, Feb 26, 2026). Explains DataCite–ORCID auto-update and credits creation of a ‘PID Graph’ encompassing datasets, code, etc. (info.orcid.org).
- Crossref. *DataCite* (Crossref Community, 2025). Discussion of Crossref vs DataCite missions: Crossref “makes research outputs easy to find, cite, link, assess, and reuse”, DataCite “leader in persistent identifiers for research” (www.crossref.org). Emphasizes accessibility of metadata via APIs (www.crossref.org).
- Angela Dappert et al. *Connecting the Persistent Identifier Ecosystem: Building the Technical and Human Infrastructure for Open Research* (Data Science Journal 2017): Comprehensive review of PIDs including ORCID, Crossref, DataCite. Key insight: “ORCID focuses on researchers, Crossref on articles, and DataCite on data” (datascience.codata.org). Contains user story diagrams showing linking (see Figures 1–4, e.g. researcher linking data to ORCID (datascience.codata.org)).
- ORCID users’ Google Group. *ORCID sync?* thread (Mar 13, 2023). Notes OpenAlex currently uses ORCID public API (no auto-update) and suggests using Crossref/DataCite auto-update as a sync path (groups.google.com).

- DataCite. *DataCite Metadata Is Now Integrated in OpenAlex* (DataCite Blog, Oct 9, 2025). Announces integration: “92 million DataCite DOIs now available in OpenAlex”; expects this to increase discoverability of datasets, preprints, software (datacite.org). Quotes OpenAlex and DataCite staff on the significance.
- OpenAlex. *Do you index DataCite DOIs?* (OpenAlex Help, Nov 7, 2024). As of late 2024, 6.4M DataCite works were in OpenAlex, with a rewrite underway to fully ingest all DataCite records (target early 2025) (help.openalex.org).
- OpenAlex. *Works Overview* and *Expansion corpus* (OpenAlex Help Center; checked 20 October 2026). The default core catalog contains over **320 million works**; core and expansion together contain over **510 million works**. Works come from Crossref, MAG, DataCite, PubMed, HAL, repositories, and other sources (Works Overview; corpus definitions).
- University of Cambridge. *ORCID IDs in Research Data Management at Cambridge*. Case study: Cambridge uses ORCID in its Apollo repository; when a researcher submits data (with ORCID), Apollo mints a DataCite DOI and DataCite auto-updates ORCID record, linking researcher to dataset and vice versa (ukorcidsupport.jisc.ac.uk).
- OpenAlex. *How does OpenAlex work?* (OpenAlex Help). States OpenAlex “pulls information” from agencies like Crossref/DataCite and repos, then matches to PID entities (ROR, ORCID, ISSN) to build the knowledge graph (help.openalex.org). Lists core sources (Crossref, DataCite, PubMed, HAL, ORCID, arXiv, etc.) being ingested (help.openalex.org).
- Crossref. *Highlights of a very busy year: our 2025 annual report* (Crossref Blog Dec 18, 2025). Reports Crossref’s coverage “across all **180 million records**” in the metadata store (www.crossref.org).
- Crossref. *Joint statement on research data* (Crossref Blog Nov 28, 2023). Crossref/DataCite note that nearly 10 million data citations tracked, and emphasize data availability practices. Earlier, *Data citation: let’s do this* (Oct 2018) highlighted growth in journal articles citing data, with Crossref and DataCite collaborating on standards (www.crossref.org).
- ORCID. *10M ORCID iDs!* (ORCID Blog, Nov 20, 2020). Celebrate milestone of **10 million ORCID iDs** issued, up from 1.5M in 2015 (info.orcid.org).
- G. Bilder. *DOIs unambiguously and persistently identify published…?* (rogue-scholar.org, Sept 2013). Notes that Crossref DOIs have vastly improved interoperability of citations, but also warns that DOIs from different Registration Agencies (e.g. DataCite vs Crossref) are only interoperable to the extent their constituencies overlap (rogue-scholar.org).
- J. Haunschild et al. (2026). *The OpenAlex database in review: Evaluating its applications, capabilities, and limitations.* (Journal of Informetrics). Highlights that OpenAlex is proving strategic across the globe (esp. in Global South) but also flags issues in data quality (affiliation, language, doc type) and advises validation (www.sciencedirect.com). (Open access)
- A. Scholefield. *ORCID IDs in Research Data Management at Cambridge* (UK ORCID Support). Summary of Cambridge case study (ukorcidsupport.jisc.ac.uk).
- **Additional references:** Crossref community blogs on metadata matching and integration (www.crossref.org), ORCID integration documentation, DataCite/ORCID support pages on metadata linking; various scientometric comparisons of databases and citation networks (datascience.codata.org) (www.sciencedirect.com) (datascience.codata.org).

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://www.crossref.org/blog/highlights-of-a-very-busy-year-our-2025-annual-report
2. https://datacite.org/blog/datacite-metadata-is-now-integrated-in-openalex
3. https://info.orcid.org/10m-orcid-ids
4. https://help.openalex.org/data/works
5. https://help.openalex.org/data/works/corpus
6. https://www.crossref.org/news/2015-10-26-orcid-launches-crossref-and-datacite-auto-update
7. https://ukorcidsupport.jisc.ac.uk/guidance/case-studies/using-orcid-ids-in-research-data-management-at-cambridge
8. https://help.openalex.org/hc/en-us/articles/28932712154391-How-does-OpenAlex-work
9. https://datascience.codata.org/articles/700
10. https://www.crossref.org/categories/datacite
11. https://www.sciencedirect.com/science/article/pii/S1751157726000337?dgcid=rss_sd_all
12. https://info.orcid.org/auto-update-has-arrived-orcid-records-move-to-the-next-level
13. https://help.openalex.org/hc/en-us/articles/24347019383191-Where-do-works-in-OpenAlex-come-from
14. https://rogue-scholar.org/records/34rxq-wc797
15. https://www.crossref.org/community/datacite
16. https://www.production.crossref.org/categories/linking
17. https://www.sciencedirect.com/science/article/pii/S1751157724001305
18. https://www.crossref.org/deprecated
19. https://ror.readme.io/docs/funder-registry
20. https://www.crossref.org/blog/strengthening-support-for-data-citations-and-saying-goodbye-to-event-data
21. https://info.orcid.org/orcid-and-datacite-supercharge-your-research-visibility
22. https://support.datacite.org/docs/consuming-citations-and-references
23. https://support.datacite.org/docs/eventdata-guide
24. https://info.orcid.org/ufaqs/can-i-create-records-for-my-researchers
25. https://www.cambridge.org/core/journals/international-organization/information/author-instructions/submitting-your-materials
26. https://help.openalex.org/data/authors/orcid
27. https://support.orcid.org/hc/en-us/articles/360006972413-Does-an-ORCID-iD-assure-my-identity
28. https://doi.org/10.6084/m9.figshare.2008206
29. https://help.openalex.org/access/example-costs
30. https://help.openalex.org/access/pricing
31. https://help.openalex.org/api/filtering
32. https://help.openalex.org/hc/en-us/articles/27629361012119-Do-you-index-DataCite-DOIs
33. https://help.openalex.org/hc/en-us/articles/27714298573719-Fix-errors-in-OpenAlex
34. https://groups.google.com/g/openalex-users/c/7A2QC1O5miE

35. https://www.crossref.org/documentation/retrieve-metadata/data-citations
36. https://arxiv.org/abs/2404.02732
37. https://credit.niso.org/contributor-roles-defined
38. https://info.orcid.org/orcid-supports-the-interoperable-exchange-of-datasets

THE PUBLISHER

About IntuitionLabs

IntuitionLabs is an AI consulting, custom software development and data engineering firm serving pharmaceutical, biotechnology, medical-device and other life-science organizations. We work with clinical, regulatory, medical-affairs, commercial, quality and IT teams to connect technology decisions with the work people need to accomplish.

AI consulting and adoption

Our AI enablement services cover readiness assessments, use-case selection, governance and policies, team workshops, adoption measurement and ongoing advisory support. We help organizations structure the information layer behind AI: source material, context, permissions and maintained knowledge that make generated answers useful and reviewable. Private LLM inference and hosted AI options support teams evaluating how to operate AI with appropriate control over their data and infrastructure.

Software, data and life-science workflows

IntuitionLabs develops custom software for pharma and biotech, integrates enterprise systems, and builds data engineering and business intelligence solutions. Areas of focus include AI agents, regulatory research, medical writing, medical affairs, CMC information, competitive intelligence and clinical-document workflows. Our eTMF intelligence work includes cross-system reconciliation and inspection-readiness support.

Enterprise platforms and regulated delivery

We provide Veeva services, application support, managed services, integrations and custom applications, alongside enterprise content work involving platforms such as Egnyte. For regulated workflows, our services include GxP enablement, computer-system validation and software development addressing 21 CFR Part 11 requirements. The applicable controls, validation responsibilities and acceptance criteria are defined for each engagement.

Work with IntuitionLabs

Explore AI enablement, pharma and biotech software development, data engineering and BI, and Veeva services. Contact IntuitionLabs to discuss your workflow, information sources and implementation needs.

IntuitionLabs publishes educational research to help life-science teams make informed technology decisions. Coverage of a product or organization does not imply a client relationship, endorsement or partnership.

Website: https://intuitionlabs.ai

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.