The Publish Trust Framework Registration on OIXNET

Building a Trust Measure for Attribute Exchange

October 1, 2013

The Open Identity Exchange | APA Trust Lab

By Jennifer Golbeck, Gary VandenBos, Hal Warren, Eva Winer, Bryan Dennis, and Debra Diener

EXECUTIVE SUMMARY

"Open and machine readable" is becoming the new default for publishing data. Ubiquitous computing and increasing volumes, variety and velocity of data create a growing need for a modern digital infrastructure for secure exchange and reuse of information.

Data repositories and **data registries** provide a technical foundation and a number of benefits for optimal discovery, validation, provenance tracking, and interoperability of data. But currently there is no standard practice for measuring trustworthiness of the attributes being exchanged online.

As a solution to this problem, a rapidly emerging ecosystem of **trust frameworks** provides a set of business, legal, technical and privacy standards and tools for building trust in online identity across different sectors, including the domain of scientific communication.

To manage the complexity of **online trust** and improve interoperability among identity federations in commercial, non-profit, and public sectors, the **Open Identity Exchange** (www.openidentyexchange.org) created OIXnet, an authoritative global registry for online trust. The listings in the **OIXnet registry** will document in a transparent and a machine-readable way how various trust frameworks operate, where the data originates, and how it can be used.

This paper illustrates the benefits of a **Linked Data approach** to entity registration in OIXnet and presents the **Publish Trust Framework (PTF)** pilot (www.publishtrust.org) developed by the **American Psychological Association (APA)** to enable secure exchange, aggregation and validation of scholarly identity attributes and facilitate interdisciplinary collaboration in scientific communities.

The APA pilot uses **Semantic Web technologies**, recently named by Gartner as one of the top ten strategic and disruptive trends for the near future, to leverage "OIXnet trusted" registration, develop reliable mechanisms for verifying scholarly identity, and backing authorship claims. The **APA Trust Factor** is computed on a core set of machine, environment, human and attribute variables and integrates data from attribute providers into a single score.

Adopting **VIVO** (www.vivoweb.org) as an open source **semantic platform** in the pilot links together major research institutions, government agencies, data centers, scholarly publishers and societies, and allows for augmentation of online assertions with provenance information, trustmarks and trust values for online consumption, interchange and analysis across social networks within and beyond the academic community.

Table of Contents

1. The Challenge of Online Trust

Trust has become a key to success in building identity online. Millions of people exchange information via social networks, blogs, online forums, professional communications and hundreds of other venues. However, there are few reliable and transparent mechanisms in place for verifying who they are and to support the trustworthiness of this information.

A rapidly emerging online ecosystem of **trust frameworks** provides a new infrastructure for backing attributes and claims by a credible third party, as well as a set of business, legal, technical and privacy standards, and tools for building trust in online identity across different sectors, including the domain of scientific communication.

Building Trusted Online Identities in Science

The **Publish Trust Framework (PTF)** pilot was developed by the **American Psychological Association (APA**) in partnership with the **Human-Computer Interaction Lab** at the University of Maryland and the **Open Identity Exchange (OIX)** Trust Framework Provider, to improve online trust in scholarly identity and scientific attribution. Essentially, the PTF uses Semantic Web technologies to augment online assertions with provenance information and a trust metric to show they were issued by a trustworthy source.

Creating Communities of Practice

Adding trust values to online identities and the ability to propagate trust across existing research networks with interoperable communication technologies used in the PTF make it possible for scientists and researchers to move well beyond their "core" network of colleagues and collaborators to a wider world of potential collaborators around the globe.

APA will produce a Trust Factor based on the PTF registration variables and publish the score with the PTF VIVO record. With a baseline for trustworthiness of online assertions, stronger communities can result. The benefit to scientists and researchers is the opportunity to create more comprehensive trusted "communities of practice" centered around research areas, interdisciplinary practice, and applied clinical contexts.

A Trust Algorithm for Attribute Exchange

The **APA Trust Factor** provides machine-computed trust values for individuals, organizations, and works across social networks. It is grounded in the National Institute of Standards and Technology

Levels of Assurance (NIST LOA) and the InCommon Federation's identity assurance criteria. It relies upon the aggregation of verified attributes of identity of a core set of variables comprised of machine, environment, human and machine networks, and policy measures used to evaluate the provenance of attributes.

For the purposes of the Trust Factor, the concept of trust, a complex and multivalent concept that means many different things within professional studies (e.g., social psychology, sociology, philosophy, economics, and computer science), will be simplified so that specific elements of trust can be measured for computational purposes.

Sociological studies generally agree that trust contains two elements: a belief and a risk. A definition that captures most common ideas is as follows: "Trust in a person is a commitment to an action based on a belief that the future actions of that person will lead to a good outcome."

In the context of the Trust Factor, users want assurance that the attributes they are viewing are true. They support their belief through the Trust Factor score next to each attribute, and can then use that to support taking actions based on this belief.

The Trust Factor considers three major categories of attribute provider traits. First, it measures confidence in identity as measured by NIST LOA. Second, it considers several factors that measure the accuracy of shared attributes. This includes how attributes are verified, what is presented, and policies surrounding errors and correction. Finally, it measures policies that protect attribute owners' rights, including opt-in or opt-out for data sharing and other privacy-related issues.

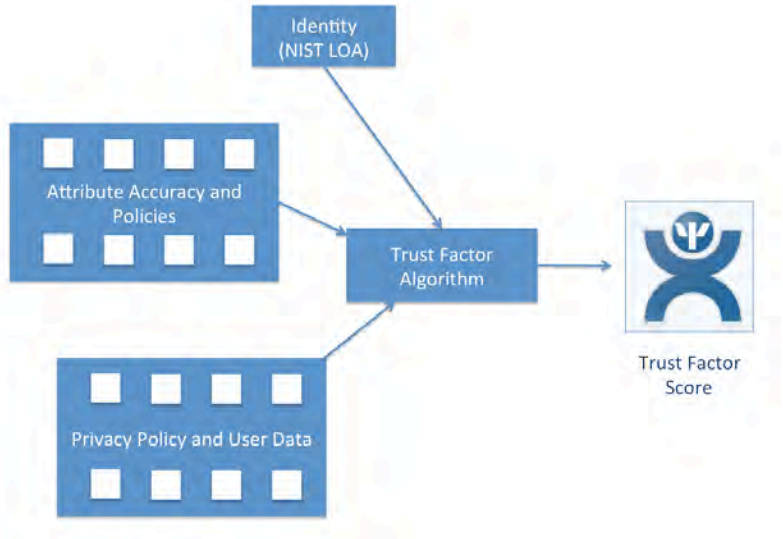

Figure 1. Traits considered in computation of the Trust Factor Score.

The final algorithm is currently in development. It is built on a combination of the measures for each of the three traits described

Users want assurance that the attributes they are viewing are true.

above. Experts will evaluate candidate methods for integrating data from attribute providers into a single score. Their input will help refine the algorithm's parameters so that the Trust Factor scores closely match the expectations of users.

2. Leveraging Trusted Data with OIXnet Registration

Managing the complexity of online trust requires a central tool for discovery, sharing and reuse of trusted identity metadata. To respond to this growing need for interoperability among identity federations in commercial, non-profit, and public sectors, the Open Identity Exchange is building OIXnet, an authoritative global registry for online trust.

In support of the OIXnet goals, the Publish Trust Framework listing in the OIXnet registry will contribute to a secure exchange, aggregation and validation of scholarly identity attributes in the open identity ecosystem to facilitate expertise discovery and interdisciplinary collaboration.

OIXnet is being developed as a technology-agnostic metadata listing service and a dynamic API discovery registry that will document in a public, transparent, and machine-readable way how various trust frameworks operate, where their data originates, and how it can be used.

As the pool of trust frameworks in the registry grows, the usefulness of syntax-based queries to find matching listings for the desired business logic may decrease. To fully realize the benefits of a discovery tool such as OIXnet, the PTF registration is implemented as a semantic listing that provides a representation of a trust framework enriched with ontological concepts and enables advanced discovery and uses of the trust framework resources.

3. Making Trusted Assertions in the Publish Trust Framework (PTF)

The PTF pilot, developed by APA as a solution for trusted scholarly identity, uses open source semantic VIVO technology for secure attribute exchange between major universities, government and scholarly publishers and demonstrates the feasibility of the OIXnet registration approach for producing stronger trust frameworks.

VIVO is an ontology-driven Java application designed to facilitate the discovery of research expertise and enable scientific collaboration across disciplines. VIVO models researcher profiles based on the ontologies in the Integrated Semantic Framework. The core Vitro ontology, predecessor to VIVO, was originally developed at Cornell

University. VIVO harvests data from verified sources like publication databases (PubMed, Scopus, APA PsycNET), institutional human resources databases, grants, and data repositories, and ingests them into the matching researcher profiles.

Specific objectives of the Publish Trust Framework pilot are:

- Improve online trust in scientific communication through rapid attribute exchange that is credible, secure, interoperable, easy to use, and measurable
- Develop trust assessment algorithms that reflect the trustworthiness of the source
- Deploy rich, user-focused platforms, interfaces, and environments that demonstrate the value online trust has in removing barriers to collaboration in communities of interest
- Develop technical solutions that will help define industry standards for semantic and linked data networks usable by governmental, nonprofit, and commercial entities.

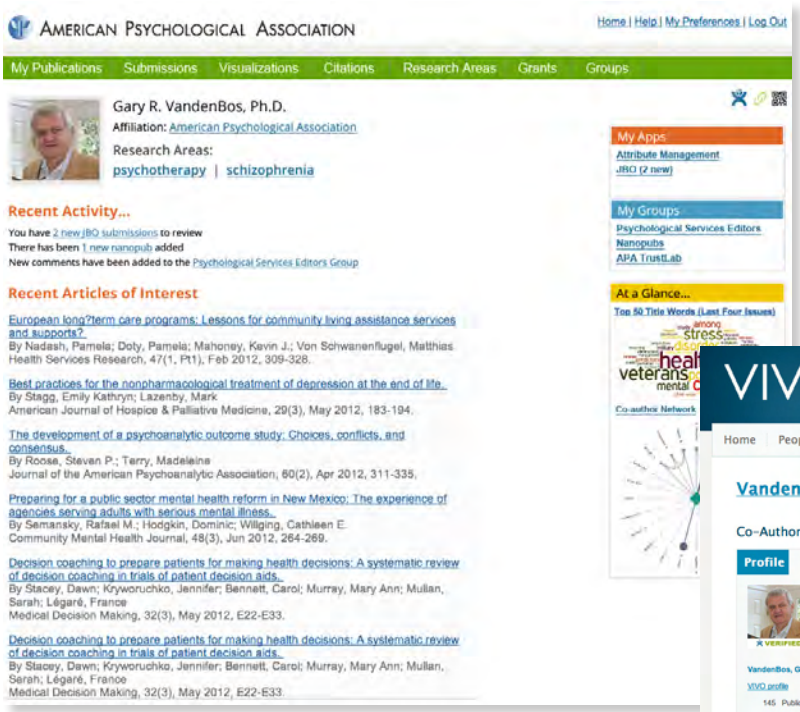

Figure 2. Example of secured author page at author.apa.org

The PTF pilot is centered on authors of articles published in APA scholarly journals, and produces publisher-validated trusted assertions of authorship, enabling scientists to claim their works, update their VIVO profiles, connect with other experts, and use new semantic tools within their journal communities for research.

Figure 3. Example of a verified VIVO profile on vivo.apa.org

The pilot uses a secure closed network of identity providers (author.apa.org) and relying parties (vivo.apa.org) to support single-sign-on between VIVO nodes, trust assertion delivery and consumption within the framework. Implementation of secure data exchange in the PTF framework leverages an enhanced SSL certificate, an RDF XML payload of core variables and a VPN to allow interoperability between nodes within the VIVO semantic network. As attributes of authorship are extended from author.apa.org to the APA VIVO profile for public consumption, the status of the individual's VIVO page changes from "unverified" to "verified".

Data in VIVO can be private or public. Public attributes can be readily exchanged with other VIVO and semantic web applications. Interoperable RDF format and Linked Data standards allow researchers to share and reuse data across various networks and external applications.

The Publish Trust Framework is a Virtual Private Network (VPN) where member institutions present Secure Socket Layer (SSL) 256 bit, enhanced verification certificates for trusted attribute exchange. A server answering at https://vivo.publishtrust.org routes attributes from member institutions to other organizational members for consumption, based on permissions set by attribute owners. These attributes are presented as URIs in VIVO RDF and JSON for consumption by trust framework member organization servers.

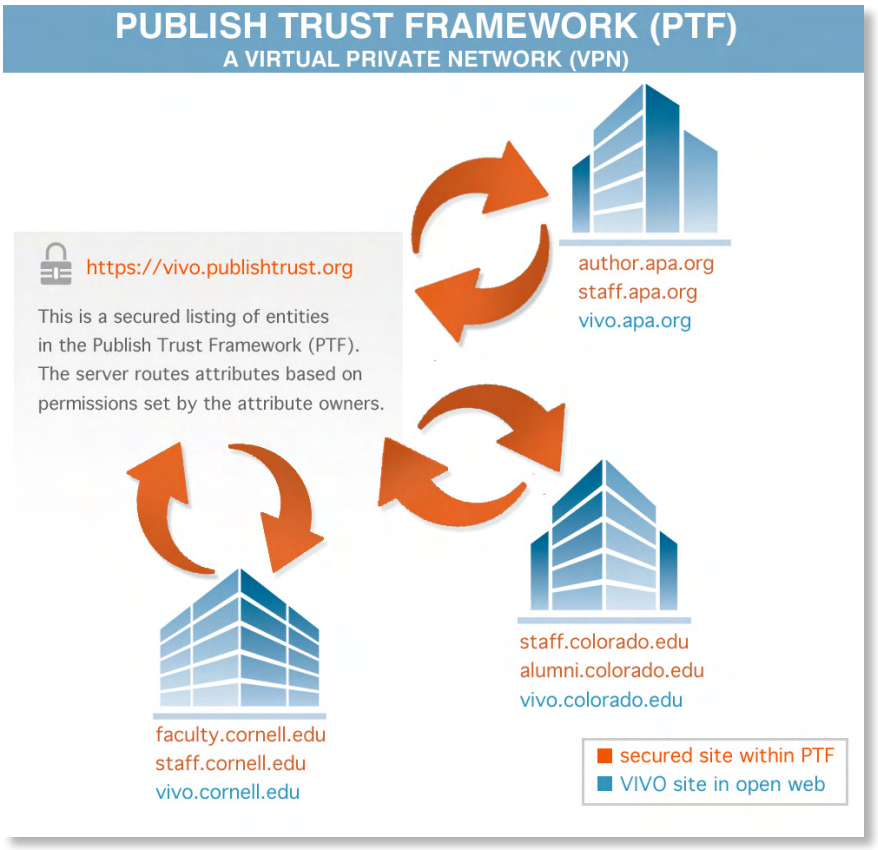

Figure 4. Publish Trust Framework

Data for a community of over 11,000 authors is now available on APA VIVO instance vivo.apa.org

Exchanges result as authorized RDF-XML payloads that can be included within other VIVO instances as Relying Parties. This is indicated by a trustmark, which links back to RDF-based metadata supporting the claim. Attribute provider credentials contain a description of the assertion including a description of the source, the relationship between the source and the account holder, and a definition of the assertion made transparent at the base URL for the attribute.

As an example, in the trust framework, authorship attributes stored at vivo.apa.org can automatically populate the author's VIVO profile at vivo.cornell.edu as a Linked Data inclusion statement. Attribute management and permissions can be set from any VIVO node in the trust framework. Attributes remain at their origin and up-to-date with owner-controlled privacy settings.

Authors remain in control of the privacy constraints for the attributes they have extended, and can retract those claims at their discretion. Claims can be challenged by other known identities in the framework, or anonymously.

The framework also provides an ability to anonymously assert expertise in a certain field. A clinical psychologist, for example, can offer advice on Facebook as a specialist on eating disorders. Anonymous ability is especially important in animal research and other sensitive research areas.

APA is currently working with Cornell University, Indiana University and the University of Colorado to source authorship attributes for works published by APA and authored by the institution's faculty.

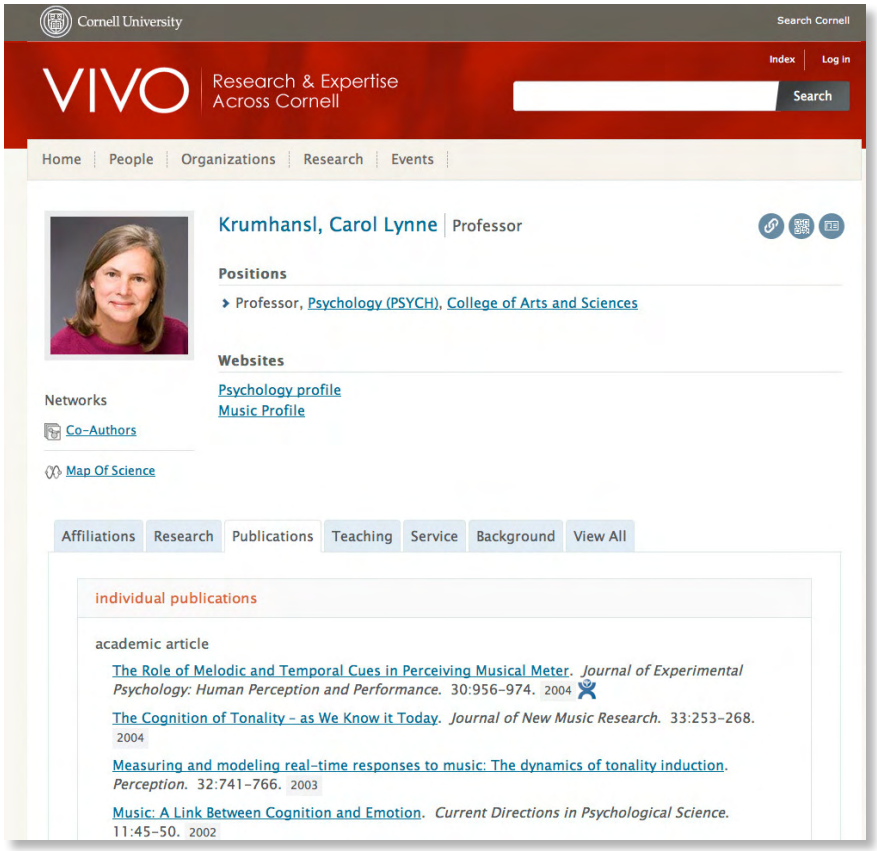

Figure 5. A sample VIVO Cornell profile with an APA trustmark next to the first publication. The trustmark contains information backed by APA and verifies authorship and journal where the article is published.

Benefits of Using the VIVO Platform as a Semantic Registry

In addition to providing an attribute exchange mechanism within VIVO, the Publish Trust Framework also requires member organizations to register basic conditions of machines, network environment, attributes, and policies. These conditions are expressed as a VIVO Record at vivo.publishtrust.org. The Publish

Cornell University is the first consumer of the trustmarked APA author attributes. Using federated authentication allows attribute exchange and linking to APA attribute servers and resources from the account holder's VIVO profile at Cornell and other member institutions.

VIVO connect share discover

Trust VIVO Server also informs the OIXnet registry for attribute and provenance consumption by other OIXnet Trust Frameworks.

Adopting VIVO as a platform for semantic registration of the PTF on OIXnet opens up an opportunity to create a machine-readable information hub for direct data exchange between applications and linking the trust framework information graph with datasets in the Linked Open Data cloud.

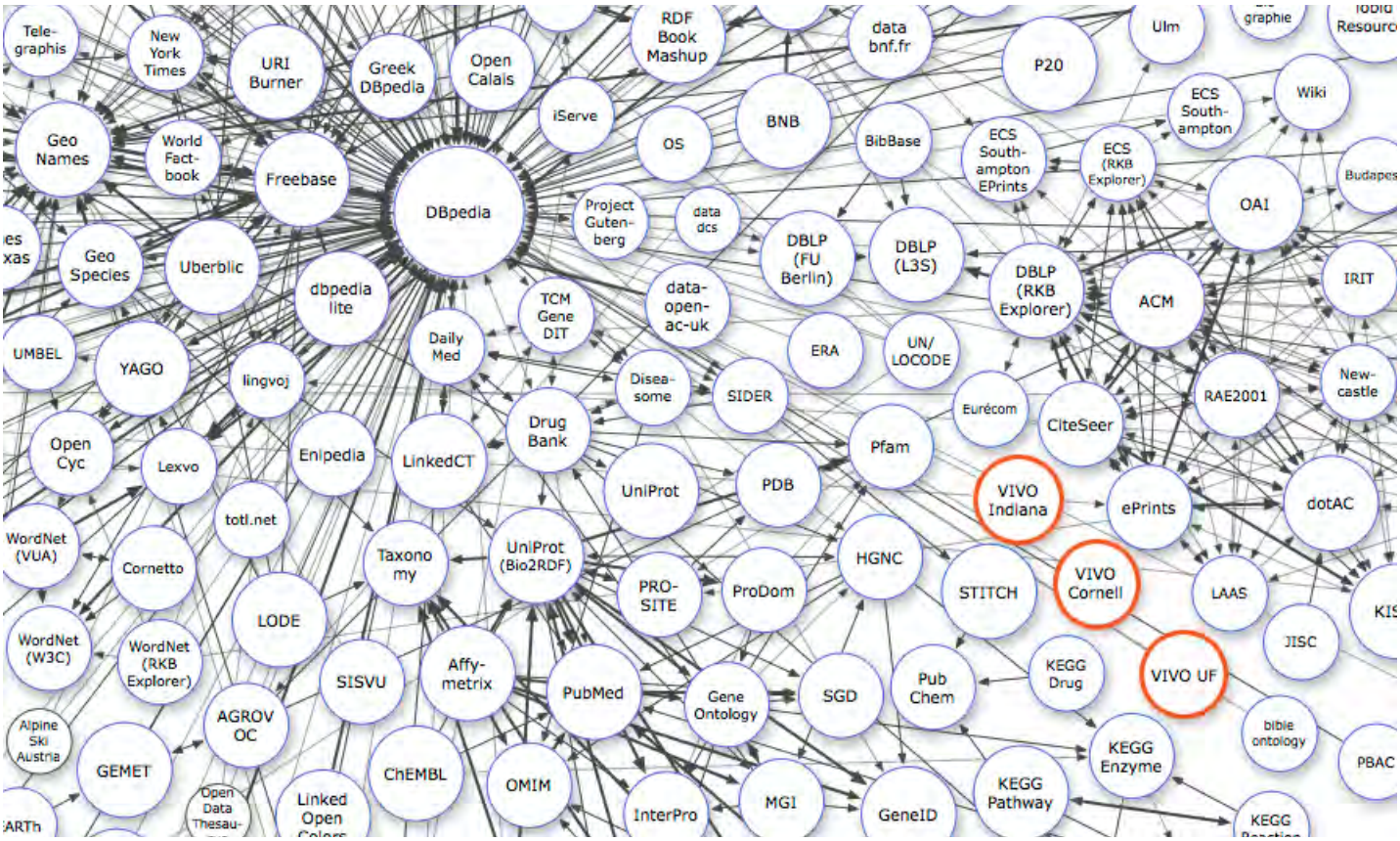

Fig. 6. VIVO data are published and accessible in the Linked Open Data cloud

VIVO Trust Framework Ontology

A lightweight Trust Framework Ontology is proposed as an extension of the core VIVO ontology to allow the Publish Trust Framework VIVO instance to represent a PTF listing in the OIXnet registry using Semantic Web standards.

The data model in the vivo.publishtrust.org deployment reflects the properties of the APA Trust Framework, including server trust, levels of assurance for identity attributes, trust algorithms, supported standards, attribute provider authority, persistent identifiers, privacy policies, assessors, and relying parties. One of the advantages of the open-source VIVO ontology is that different types of trust frameworks can customize the ontology based on the community or industry represented.

OIX represents a wide range of commercial, non-profit, and government interests from Google to Verizon, from the US to the UK. Adoption of a registry platform based on semantic standards and formatting RDF data according to a common representation format would allow users to search and aggregate data from different trust frameworks in a meaningful way and provide transparent trust framework conditions for reasoning by machines and evaluation by consumers.

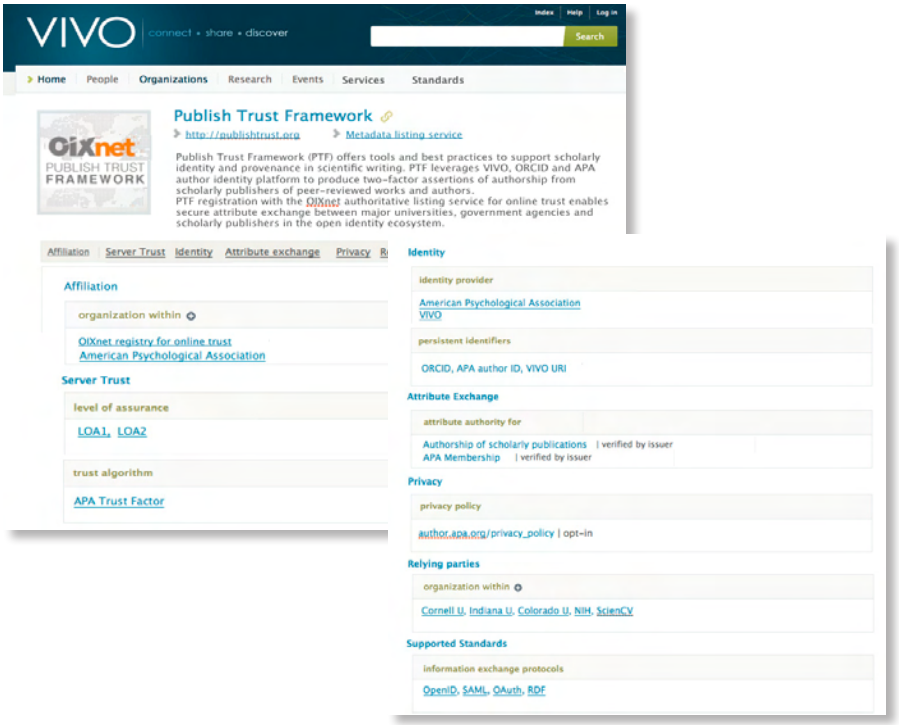

Fig. 7. PTF registry listing on vivo.publishtrust.org

Persistent Identifiers for Resource Disambiguation

In the APA Trust Framework, data is represented as RDF (where URIs identify people, organizations, events, equipment, and other entities) and published as Linked Data.

While VIVO and Linked Data tools automate the contributor-research linkage across multiple datasets, they don't fully solve the problem of author disambiguation. To clearly link authors and all their name variants with their attributes, APA developed an infrastructure for integrating Open Researcher and Contributor ID (ORCID) into the trust framework.

ORCID is a newly-minted central registry of researchers that crosses disciplines, sectors, and national boundaries with a mission of enabling reliable attribution of published works to authors and contributors.

More specifically, embedding ORCID identifiers into the Publish Trust Framework workflow allows users to:

- uniquely attach author identity attributes to research objects such as manuscript submissions, published articles, citations, experiments, patents, grant applications, professional awards, etc.
- support disambiguation of authors and reliable linking to external datasets

APA Trustmark

- facilitate the creation and maintenance of author and reviewer profiles and improve author search
- track and manage research findings by mapping URIs in different systems to ORCID
- enable ORCID as a single sign-on control to access APA author hub, VIVO and a broad range of APA resources

The ORCID identifier is part of the VIVO data model. A new VIVO ORCID Updater tool developed by APA provides automatic updating of an author's publication data in the ORCID registry from VIVO RDF data.

The Updater extracts Digital Object Identifiers (DOIs) for publications from author.apa.org profiles. Metadata associated with DOIs are extracted through a DOI lookup and pushed to ORCID.

The automatic pushing of publication data and reviewer statistics to ORCID ensures that highly proofed publication data is available in the registry, and contributes to the evolving research attribute exchange network.

Data Exchange Standards

Semantic Web technologies are based on open, secure, and effective standards defined by the W3C in order to reduce complexity in data exchange between applications.

Adoption of the VIVO ontology as a semantic data interchange standard at the national and international level offers an easy way to exchange data between institutions and applications without developing complex metadata schemas or specific APIs.

Once a researcher verifies identity and publication information in vivo.apa.org or another local instance, VIVO becomes an authoritative single source for attribute exchange with a variety of external commercial and government systems for research analytics.

A growing number of organizations, consortia and government agencies are engaging with VIVO [2] and the Publish Trust Framework to develop interoperabilities for data interchange:

- U.S. government initiatives such as STAR METRICS (starmetrics.nih.gov) and SciENCV (scienceofsciencepolicy.net/SciENCV) are focusing on research data sharing and reuse and on the analysis of the impact of federal investments in science, particularly with respect to job creation and economic growth

- The European Union (EU) euroCRIS partnership (eurocris.org) is exploring interoperation with VIVO and convergence of the VIVO semantics with the widely used EU-recommended Common European Research Information Format (CERIF) standard, as well as using VIVO interface as a front-end to the CERIF research information repository
- Collaboration with the Consortia Advancing Standards in Research Administration Information (casrai.org) aims to facilitate development of a common data dictionary and seamless research interoperability

The alignment of VIVO, ORCID, CERIF, and CASRAI standards will advance a common technology-neutral approach to data exchange and reuse between research teams, institutions, and funding agencies on a global scale.

4. Key Privacy Policy Enhancing Requirements for Attribute Sharing in the Publish Trust Framework

The PTF is privacy enhancing and promotes privacy in multiple ways. This is done by:

- drawing on authoritative sources for the attributes being shared
- providing visibility of the authoritative source and giving users confidence that the information is correct, not self-reported
- providing APA authors control over their attributes and the works with which they want to be associated
- affording users a mechanism by which to challenge information and assertions about them, and
- providing the authors with the ability to update these choices whenever they wish to do so.

APA authors have control over all their attributes; they decide if and how they want to share their information.

Additionally, by using ORCID profiles, APA authors have a unique identifier by which to distinguish and disambiguate their identities. APA authors decide how they want to share their information. By using the VIVO Platform as a semantic registry, the PTF provides an environment in which APA authors can collaborate while also providing them with control over their personally identifiable information.

The value of a semantic registry is that it enables machine readable aggregation and analysis of data elements, and provides flexible data handling that can be readily adapted to meet the changing needs of the user community.

The PTF is also unique in that it combines a “closed” community of APA authors with an “open” system of cooperating institutions (i.e., APA VIVO). In using the VIVO platform, the PTF Privacy Policy has to very clearly alert authors who use it about its dual nature.

No personally identifiable information about the user accessing the work (or works) is tracked, aggregated distributed or disclosed.

How does this dual approach enhance privacy? PTF uses an "opt in" approach affording APA authors the ability to decide what attributes they choose to share on VIVO (e.g., verified authored works which are marked differently than "public record" publication data) and which attributes they wish to keep private (e.g., email address). APA authors will have the option to extend verified information (e.g., works they have claimed and APA information they have verified) to the public VIVO site.

The Publish Trust Framework's Privacy Policy details key privacy requirements

APA authors coming to the PTF website will find a Privacy Policy outlining all of the important privacy protections being accorded to them and their personal information. In addition to the above-described controls, they learn that personal information collected on the PTF will not be shared, sold, rented or released to third parties except for the purposes for which their consent has been given or without their prior express consent (except if such release is required by law).

The website is designed to provide a secure community for APA authors. Part of the website's value for the "closed PTF community" comes from the ability to provide authors with aggregate data about the frequency with which their works have been accessed. The use of "cookies" for this type of tracking allows members to make more focused discovery. There will be no personally identifiable information about the user accessing the work (or works) tracked, aggregated, distributed or disclosed. And this type of "cookie" will only be used in the ways approved and specified by the PTF "closed community" member.

The PTF Privacy Policy also contains information for "International Visitors." This type of information should be a Privacy Policy staple given the explosion of the Internet and blurring of geographical borders. In this paragraph, users are alerted that the PTF website and services are hosted in the United States but that they may be used by visitors outside the United States. International visitors from the European Union, or other regions, are advised that the laws governing data collection and use in their respective country might differ from United States law. They are further advised that they are transferring their personal data to the United States and by doing so, are consenting to that transfer.

PTF's combined privacy requirements and protections will enable APA authors to feel confident when using it. They will have maximum flexibility to select the personal attributes they wish to share; select the contexts in which they wish to share their data; and select with whom they wish to share their data. They will have this control over their personal information while also taking full advantage of other important collaborative tools like VIVO and ORCID.

5. Beyond the PTF: Linking Experts, Publications, Institutions, and Governments

The growing VIVO community is expanding across universities, data centers, medical schools, government agencies, research institutions and professional societies. There are currently over 100 VIVO instances at the leading universities in the U.S. and dozens of projects worldwide. Partner implementations include APA, USDA, EPA, the ORCID Initiative (Open Researcher and Contributor ID), euroCRIS (The European Organization for International Research Information), Symplectic Research Management for Higher Education (UK), Chinese Academy of Science, and the Australian National Data Service (ANDS) VIVO project.

The ANDS collaboration makes an interesting use of VIVO as a research data registry, aiming to produce a metadata triple store that will aggregate and brand research data sets produced by universities and other research organizations.

Enriching the OIXnet registry architecture with the Linked Data principles can help build a robust online trust ecosystem within which data attributes for individuals, institutions, and governments are exchanged, aggregated, and shared to speed research and improve communications in science.

6. Conclusions and Next Steps

The trusted online identity ecosystem is a dynamic and evolving area. We envision that leveraging Linked Data and semantic tools along with the trusted OIXnet registration will support open innovation in science and academic publishing and will help create collaborative research platforms and multidisciplinary communities of practice.

Adoption of an open source VIVO platform based on semantic standards and RDF common representation format would make it possible to search and aggregate data from different trust frameworks in a meaningful way and provide transparent trust framework conditions for machine analytics and human consumption.

Development of the Trust Factor algorithm has proceeded with careful consideration of security, reliability, and use expectation. As the algorithm is finalized, future work will focus on collecting data from a wide base of attribute providers and validating and adjusting the scores to meet with expectations.

We will also work on developing auditing procedures to validate the processes that attribute providers claim to have in place. This will ensure that the Trust Factor scores, which rely on self-reported details about procedures, will accurately reflect the reality of their process.

References

1. Gartner Identifies Top Technology Trends Impacting Information Infrastructure in 2013. Press Release March 6, 2013. http://www.gartner.com/newsroom/id/2359715

2. Börner, K., Conlon, M., Corson-Rikert, J., & Ding Y. (2012). VIVO: A semantic approach to scholarly networking and discovery. San Rafael, Calif.: Morgan & Claypool.

3. Kuter, U., and Golbeck, J. (2010). Using Probabilistic Confidence Models for Trust Inference in Web-Based Social Networks, ACM Transactions on Internet Technology, 10(2), pp. 1-23

4. Wood, D., Zaidman, M., Ruth, L., & Hausenblas, M. (2013). Linked Data: Structured data on the Web. Manning Publications Co., 375 p.

About APA Trust Lab

The APA Trust Lab is a research and development initiative with a strategic vision for the future viability of the scientific publishing ecosystem.

The Lab continues the long tradition of innovation in knowledge dissemination at the American Psychological Association (APA). APA's research databases on PsycNET platform contain almost 180,000 articles from 79 journals published by the APA. The APA professional community includes more than 134,000 members, 116,000 authors, 78,000 reviewers and 3,500 editors/editorial board members.

Trust Lab projects focus on developing strong online identities and the ability to control how information is shared and used based on semantic web technologies that enable provenance of information and interoperability between systems.

For more information: http://trustlab.org

Contact Information

American Psychological Association
Publishing Innovation Dept.
750 First Street, NE
Washington, DC 20002-4242
innovation@apa.org
800-374-2721 or 202-336-5500
TTD/TTY: 202-336-6123