

Professional Networking and Expertise Mining for Research Collaboration

THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

<http://profiles.catalyst.harvard.edu>

Profiles Research Networking Software¹ is an open source tool to speed the process of finding researchers with specific areas of expertise for collaboration and professional networking. Profiles RNS imports and analyzes “white pages” information, publications, and other data sources to create and maintain a complete searchable library of web-based electronic CV’s. Built-in network analysis and data visualization tools allow administrators to generate research portfolios of their institution, discover connections between parts of their organization, and understand what factors influence collaboration.

Profiles RNS self-populates a database of publication history, research interests and professional relationships for each investigator in an organization. Integrated visualization and search tools make Profiles RNS easy to use, and its customizable look-and-feel allow Profiles RNS to be integrated into an existing website or set up as a stand-alone site. Profiles RNS data is also accessible through an API to power other applications.

Professional Networks

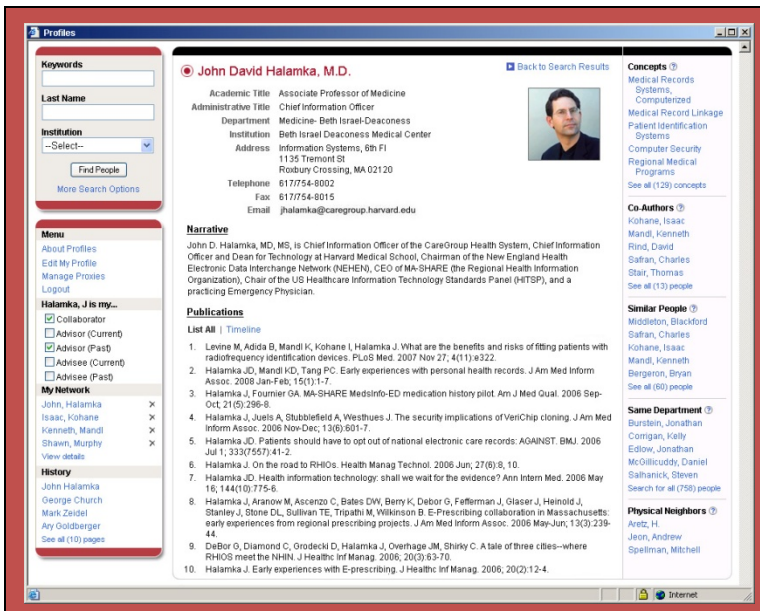
Profiles RNS pioneered the concept of “passive” and “active” networking, which not only enables the website to be a useful and exciting tool on day one, but also allows users to expand its content with information about social networks that only they know.

- Passive networks are *automatically* created based on current or past co-authorship history, organizational relationships and geographic proximity. It extends these networks by discovering new connections, such as identifying “similar people” who share related keywords. Offering these additional suggestions, Profiles RNS can lead you to unexpected opportunities for collaboration and new sources of expertise.
- Users can *manually* create active networks by identifying advisor, mentor and collaborator relationships with colleagues. Profiles RNS will soon support the OpenSocial standard, which will let researchers use the same types of plug-in collaboration gadgets found on LinkedIn and Google within their active networks.

More Than Simple Search Results

Profiles RNS provides much more useful information than typical directory listings or ordinary literature searches. Algorithms analyze publication data to define a researcher’s professional interests with a set of prioritized keywords. The factors used to rank and weight the significance of a specific keyword as a useful descriptor of a researcher include:

- The researcher’s position in the author list of a publication;
- The importance of a keyword as a publication topic;
- The date of a specific publication;
- The overall commonness of a keyword in the literature;
- The impact of a publication using citation information.



Profiles RNS creates a “career snapshot” that combines directory information, user-contributed content, and publications that are extracted from PubMed by the Profiles RNS Disambiguation Engine. On the right side-bar, Profiles RNS automatically identifies networks of related people and concepts (“passive” networking). On the left, users can create “active networks” with specific colleagues.

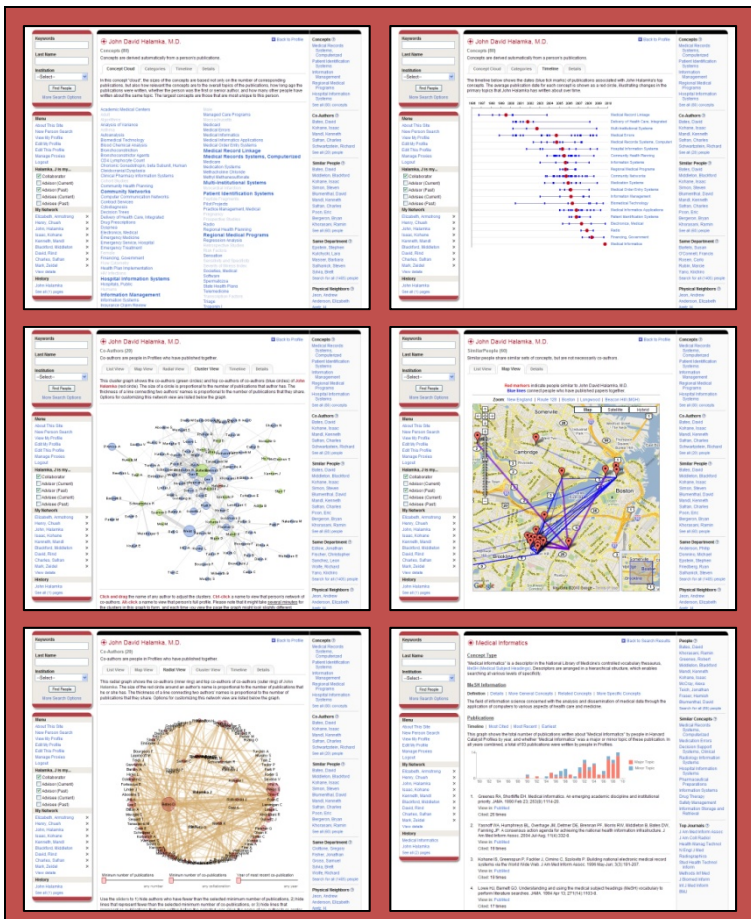
The Profiles RNS Author Disambiguation Engine

Profiles RNS uses sophisticated multi-factorial matching algorithms to build a publication history automatically for each researcher in an institution. This “Disambiguation Engine” self-populates the individual researcher overviews in Profiles RNS, and identifies the specific keywords that characterize each researcher.

Using identity information from a managed data source such as a Human Resources database, the Disambiguation Engine extracts citations from PubMed and assigns publications to specific individuals. The Disambiguation Engine uses a number of factors to build each publication history, including:

- Name permutations (e.g., first name vs first initial);
- Email address;
- Institution affiliations;
- Known co-authors;
- Journal titles and subject areas;
- Known relevant keywords.

Users (or their proxies) can add any missing publications by doing a PubMed search from within Profiles RNS or manually entering publications that do not exist in PubMed. The Disambiguation Engine learns from these changes to improve the results of the next literature analysis and update.



Various ways of visualizing networks within Profiles RNS include (clockwise from upper left) concept “clouds” that emphasize keywords that make person’s research unique; network timelines that show how those keywords have changed over time; geographic maps that illustrate where similar people are located; concept profiles that create a view of an institution’s strengths; radial graphs that present “ego-centric” views of collaborations; and co-author cluster graphs that reveal how teams interact.

Network Analysis and Visualization

An aspect of Profiles RNS that distinguishes it from other research collaboration tools is its built-in ability to analyze and visualize networks in novel ways. Centrality calculations commonly used in bibliometric and social network analysis, such as “degree” and “closeness”, determine a person, department, or institution’s position within the research community. Profiles RNS uses this to supplement traditional measures of academic success like number of publications and citation counts.

User Managed Content and Privacy Controls

Each researcher at an institution has control over her or his information. While contact details and other directory information are managed through the original source system at the institution, each user can select which sections of their overview page is displayed or hidden. In addition to editing their publication lists, users can add a photograph, short narrative summary, and awards to their profiles. To make managing researcher information easier, Profiles RNS allows proxies to be designated for each user. Proxy access is configurable, controlling the ability to edit or show/hide specific categories of information.

Extending Functionality Through an Ontology

Profiles RNS is an ontology-based Semantic Web application, meaning users with administrative privileges can extend the types of data that are supported by simply describing the new entities and their relationships in configuration files—the software code does not have to be modified. Profiles RNS comes with the VIVO ontology, an NIH funded standard for sharing information about researchers.

Federated Search and Standards-Based APIs

Profiles RNS uses standards-based web service APIs that can communicate with other computer systems through XML and RDF using Linked Open Data (LOD). This enables sites using Profiles RNS to participate multi-institution networks, such as Direct2Experts, VIVOsearch, and CTSASearch. These enable users to find potential collaborators at sites using Profiles RNS or other professional networking tools such as VIVO, CAP, Loki, or Digital Vita.

Open Research Networking Gadgets

Install the optional Open Research Networking Gadgets (ORNG) extension to plug in free third party or custom applications tailored to your institution. ORNG is based on Open Social standards and leverages the VIVO ontology in Profiles RNS. Build your own custom gadgets, or install available add-ons from the ORNG.info App Library.

Integration with ORCID

ORCID.org is a non-profit organization, which creates unique identifiers for researchers to solve the name disambiguation problem. Profiles RNS automatically creates ORCID iDs for users and synchs publications and other profile content.

Open Source Community

Profiles Research Networking Software is available for free under a standard BSD Open Source license. Over 200 institutions, including major academic health centers such as Harvard and UCSF, are members of the Profiles RNS Users Group, which holds monthly webinars to keep institutions informed of upcoming features and solicit ideas for new functionality.

Technical Requirements

- Microsoft SQL Server 2005+ database server
- Microsoft Windows IIS web server with .NET 3.5

Commercial Support



Symplectic (<http://symplectic.co.uk/services/profiles-rns>) is a software company who specializes in developing, implementing and integrating research information systems for universities and research institutions around the world. They help institutions with installation, configuration, and on-going support of Profiles RNS. Their services for Profiles RNS include:

- Installation. Install Profiles RNS locally-hosted or in the cloud.
- Data Import/Migration. Integrate Profiles RNS with institutional bibliographic and faculty systems.
- Configuration and Branding. Configure and brand Profiles RNS to meet institutional requirements.
- Training and Support. Provide training to allow for easy internal management of Profiles RNS, with on-going support available via our support site.

¹Profiles Research Networking Software was developed under the supervision of Griffin M Weber, MD, PhD, (weber@hms.harvard.edu) with support from Grant Numbers 1 UL1 RR025758-01 and 1UL1 TR001102-01 to Harvard Catalyst, The Harvard Clinical and Translational Science Center, from the National Institutes of Health and support from Harvard University and its affiliated academic healthcare centers.