

Lessons Learned from Developing RIALTO: A Research Intelligence System

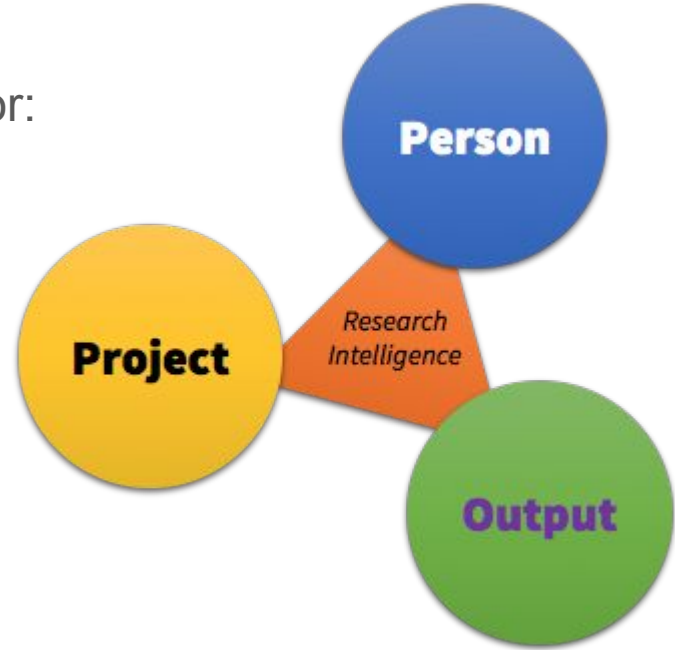
November 18 2020
OPERA Conference



What is “Research Intelligence”?

Measuring, linking and analyzing research output for:

- Compliance Reporting
- Return on “investment” (or impact)
- Collaboration analysis
- Strategic planning
- Trend analysis



What we know



WHO



HOW



WHAT

Sort of...

The What

Different systems track different kinds of outputs to varying degrees of completeness, but most are not integrated:

- Books published
- Publications published
- Students graduated
- Courses taught
- Patents filed
- Others

Current public facing Profiles systems

Stanford | PROFILES

SIGN IN EDIT MY PROFILE

Switch to Stanford-only View

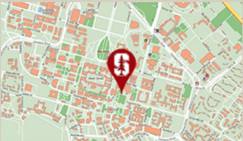
Search or Browse Profiles

Search by name, email, work phone, SUNet ID, topic or organization...

SCHOOLS INSTITUTES ADMINISTRATION

ABOUT STANFORD PROFILES


Search for faculty, staff, students and postdocs in one place to find information (Expertise, Bio, Research, Publications, and more) on people you would like to collaborate with and contact information to be able to communicate with them directly. Search for Staff to also find information on current roles, professional experience, and project work. Browse a list of people affiliated with an organization by selecting from the available menu options or search for a department or organization by name.

STANFORD CAMPUS MAP

Search the Stanford Map

RELATED RESOURCES

[Search Stanford Web](#)
[Find an Adult-care doctor](#)
[Find a Pediatrician or pediatric specialist](#)
[Tree View of University Org Code Hierarchy](#)
[Organization Manager](#)

 CAP POWERED BY CAP NETWORK

Report Accessibility Issues


Stanford | PROFILES

Search by name, email, work phone...

BROWSE SIGN IN EDIT MY PROFILE

PRINT PROFILE EMAIL PROFILE

VIEW STANFORD-ONLY PROFILE

**Persis Drell**

PROVOST, JAMES AND ANNA MARIE SPILKER PROFESSOR AND PROFESSOR IN THE SCHOOL OF ENGINEERING, PROFESSOR OF MATERIALS SCIENCE AND ENGINEERING AND PROFESSOR OF PHYSICS

BIO

TEACHING

PUBLICATIONS

Bio

Persis Drell, Provost

Drell is a physicist who has served on the Stanford faculty since 2002. She is the James and Anna Marie Spilker Professor in the School of Engineering, a professor of materials science and engineering, and a professor of physics. She is the former dean of the Stanford School of Engineering and the former director of the U.S. Department of Energy's SLAC National Accelerator Laboratory at Stanford.

Drell received her bachelor's degree in mathematics and physics from Wellesley College in 1977, followed by a PhD in atomic physics from the University of California, Berkeley, in 1983. She then switched to high-energy experimental physics and worked as a postdoctoral scientist at the Lawrence Berkeley National Laboratory. She joined the physics faculty at Cornell University in 1988.

In 2002, Drell joined the Stanford faculty as a professor and director of research at SLAC. In her early years at SLAC, she worked on the construction of the Fermi Gamma-ray Space Telescope. In 2005, she became SLAC's deputy director and was named director two years later. She led the 1,600-employee SLAC National Accelerator Laboratory until 2012. Drell is credited with helping broaden the focus of the laboratory, increasing collaborations between SLAC and the main Stanford campus, and overseeing transformational projects.

During Drell's tenure as director, SLAC transitioned from being a laboratory dedicated primarily to

CONTACT

Academic

University - Faculty
Department: Provost's Office
Position: Provost
BLDG. 10
Stanford, California 94305

University - Faculty
Department: Materials Science and Engineering
Position: Professor
University - Faculty
Department: Physics
Position: Professor

Alternate Contact

Nancy Crow
Executive Assistant to the Dean
ncrow@stanford.edu

ADDITIONAL INFO

Profiles are not complete and the data that backs it are not well positioned to provide analytics or aggregation.

The current situation is a bit like a company that tracks income and records outputs, but doesn't measure the organization's decisions or strategy impact either.

Research intelligence is understanding how a research organization's decision making and strategy affects its inputs and outputs, as well as the impact the research outputs have on the subject specific domains.

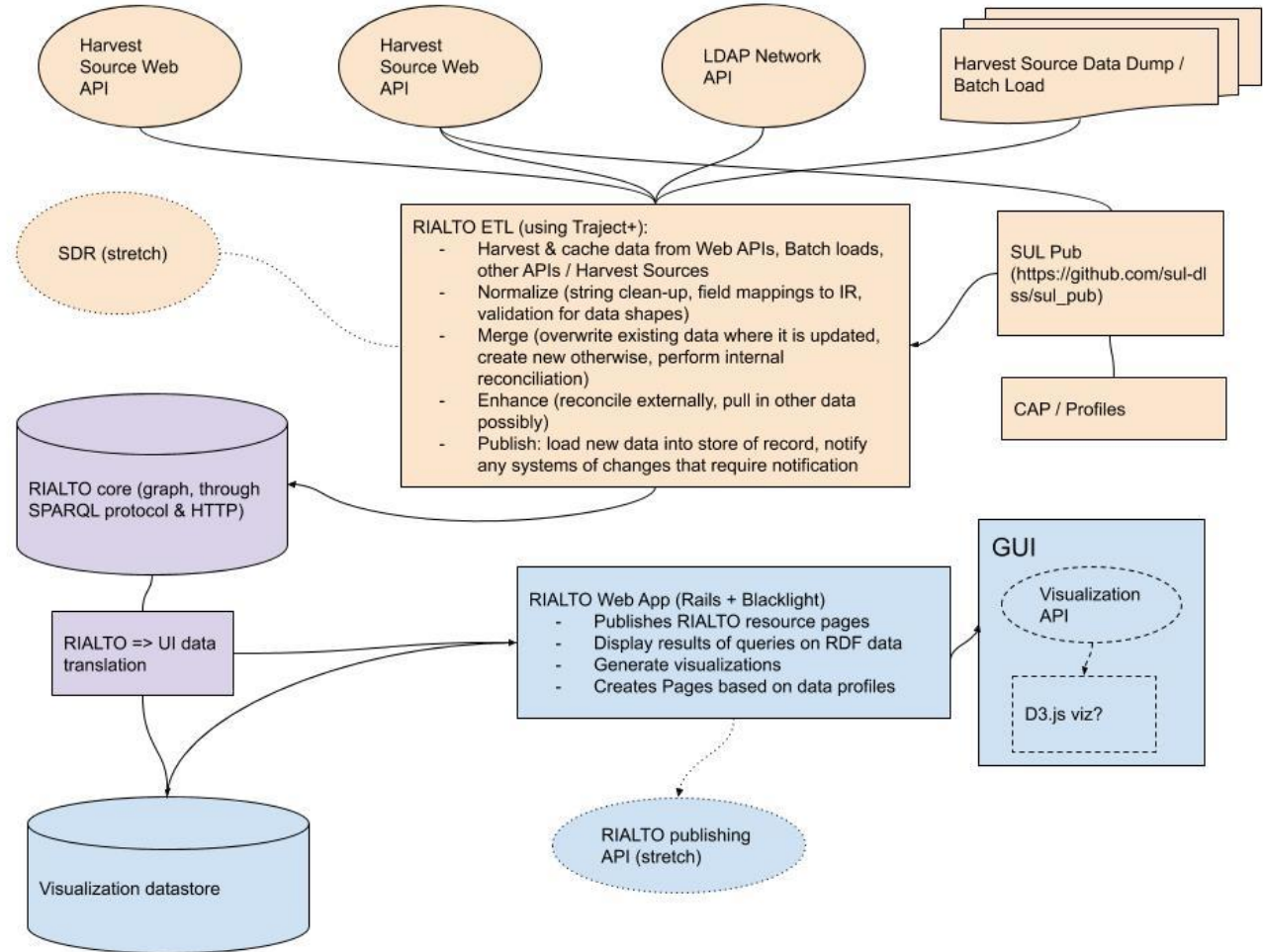
Research intelligence is an expanding domain

- Clarivate (InCites, Converis, Web of Science)
- Digital Science (Dimensions, Symplectic Elements)
- Elsevier (Pure, Scopus, SciVal)
- Open stack (VIVO)
- Custom solutions (what we did!)

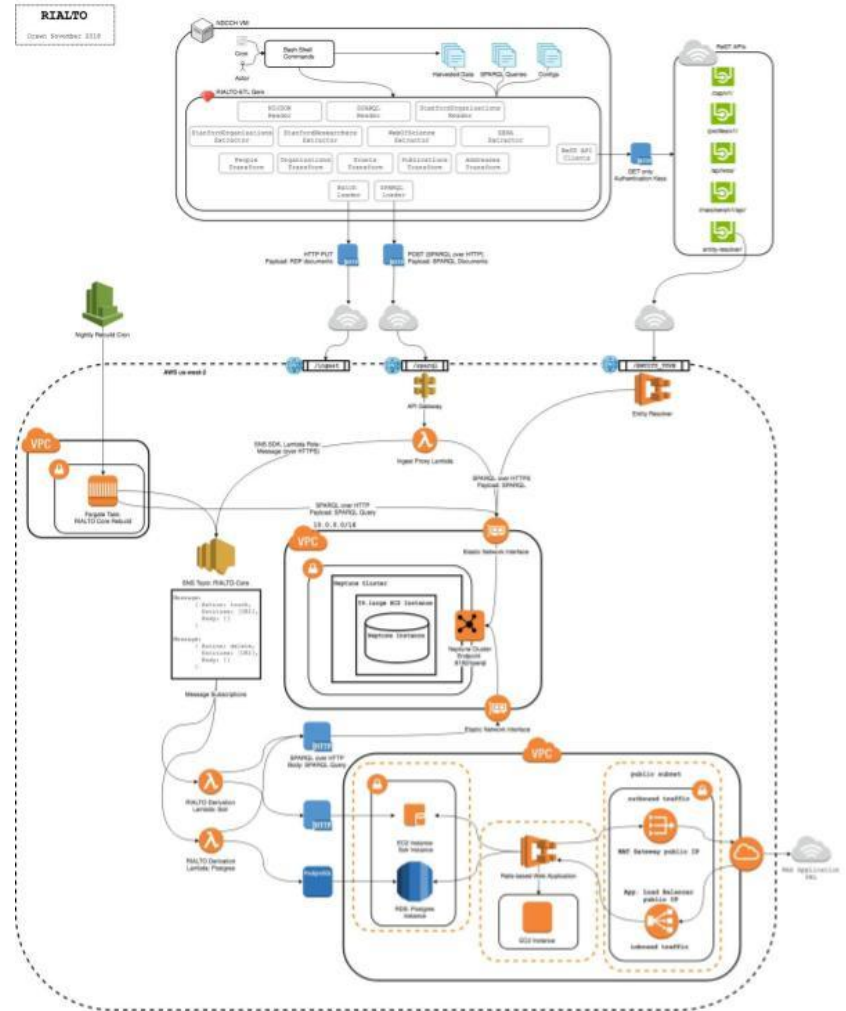
Top Use Cases We Worked On

- As the Director of the Office of International Affairs, I need to be able to map Stanford faculty members' international co-authors by institution so that I can make strategic decisions about potential university-wide partnerships.
- As a university administrator, I want to know about trends in the areas of research being produced by Stanford faculty so I can better understand where faculty interests area
- As a university administrator, I need to understand the impact of cross-disciplinary institutions on research impact so I can assess the ROI
- As a university administrator, I need to know which publications resulted from which grants so that I can fulfill reporting requirements to funding agencies

High Level Architecture



Technical Diagram



Analytics



RIALTO

[Analytics](#) [Browse](#)

Discover patterns of research, publication, and funding



Discover
Stanford
collaborations

Search



Research trends
by topic

Search



Cross-disciplinary
research output
by topic

Search

Browse

The screenshot shows a web browser window with the URL `https://rialto.dev.sul.stanford.edu/#/catalog`. The browser's address bar and tabs are visible at the top. The page header features the Stanford Libraries logo and the word "RIALTO" in large white letters on a dark red background. Below the header, there is a navigation bar with "Analytics" and "Browse" links. A search bar with the placeholder text "Search..." and a "Search" button is located on the right. On the left side, there is a vertical list of filter categories: Type, School, Department, Institute, Institution, Concept, and Country, each with a right-pointing arrow. The main content area displays the text "Previous | 1 - 10 of 128964 | Next". Below this, the first search result is for "Michael Gonzalez", listed with a bullet point. The details for this entry are: Name: Michael Gonzalez, Department: Information Technology Services (Office of Vice President for Business Affairs and Chief Financial Officer), School: Office of Vice President for Business Affairs and Chief Financial Officer, University: Stanford University, and DOI. The second search result, "I. Massa", is partially visible at the bottom.

Stanford LIBRARIES

RIALTO

Analytics Browse

Search... Search

Previous | 1 - 10 of 128964 | Next

- **Michael Gonzalez**
Name
Michael Gonzalez
Department
Information Technology Services (Office of Vice President for Business Affairs and Chief Financial Officer)
School
Office of Vice President for Business Affairs and Chief Financial Officer
University
Stanford University
DOI
- **I. Massa**

Benefits for stakeholders

- Reports built to specifically identified use cases
- Aggregation by internal only Stanford data
(e.g. department and school affiliations)
- Stanford only data available that does not exist in similar commercial systems
(e.g. non public sponsored projects)

Note, most of the data are already available to most of the stakeholders we've talked to, but getting it all, iterating, collating and then filtering by field is a lot of work!

Challenges

- Coverage gaps and incomplete data (coverage varies by subject area)
- No historical data (only current Stanford researchers represented)
- Name ambiguity leads to many false positive and false negative connections between authors, publications and grants
- Research areas are determined at the level of a journal (so all publications in a given journal will have the same research area)
- Out of date unless you constantly harvest new data
- Hard to get performance at web speed with an RDF based data store
- Architectural complexity
- System has many built in external dependencies (e.g. APIs for data)
- Data inconsistencies (usual stuff: misspellings, inconsistent use of fields, duplicate data, bad data, etc.)

Future Possibilities

- APIs, APIs, APIs
- ORCID capture and sync with ORCID profiles
- Additional data sources
- Algorithmic and AI intelligence around the linking of data

Many use cases we considered involved questions around “return on investment” based on one factor or another.

So the question is, how do you measure impact, or “return” on academic research?

Number of publications produced?

Number of grants received?

Number of citations generated?

Number of collaborators?

Altmetrics score?

Students mentored or graduated?

Courses taught?

And over what time periods?

If it takes 5 years for one important paper to become seminal in the field and start generating many citations, should this count as higher impact than a large number of more incremental papers that happen to generate a higher total number of citations over the same timeframe? If so, how do you do this? How does one algorithmically know a work is seminal?

If we build systems that get the answers to these questions wrong, we may end up making strategic decisions or incentivizing work based on flawed metrics.

Facebook measures engagement with numbers of clicks and likes, and success as ad revenue generated. Which leads to misleading and bad content, simply because it optimizes for this.

Let's avoid the same mistakes in academia by being cautious about what we optimize for when building research intelligence systems.

Having lots of data is not sufficient.

Having lots of linked data is not sufficient.

Having lots of correctly linked data is not sufficient.

Having lots of correctly linked data that is reasonably complete in scope is not sufficient.

Having lots of correctly linked data that is reasonably complete in scope with fancy reports is not sufficient.

What is first required is a good understanding of what the actual metrics are.

In conclusion...

Research intelligence requires more than aggregated data and graphs. It requires careful thought on how we interpret the data and how this interpretation is used for strategic decision making.

So what's next? Let's talk!

Peter Mangiafico

pmangiafico@stanford.edu