Knowing Ourselves: Building an Interactive Researcher Map at the University of Alberta

Luciano Frizzera,[1] Morgan Cselinacz,[2] Mihaela Ilovan,[2] Astrid Ensslin,[2] and Geoffrey Rockwell.[2]

1. Concordia University
2. University of Alberta

Introduction

> Almost all academic journals, in their mission statements, now claim to be 'interdisciplinary'; so do many academic departments (particularly English departments) and even entire universities. (Moran 2010, loc. 114)

Despite claims to interdisciplinarity, universities typically organize knowledge along disciplinary lines in departments and faculties. Inevitably, researchers and projects are constrained by such organizational structures which leads institutions to form interdisciplinary units that can accommodate alternative research configurations and initiatives that cross departments and faculties. Institutes like the Kule Institute for Advanced Study[1] (KIAS) at the University of Alberta (UofA) have been set up to encourage the development of interdisciplinary research projects, but what do we really know about the research of our colleagues and the connections among them other than their departmental affiliation? How is an institute or interdisciplinary group to know what research directions are pursued by its constituency? For that matter, how is a university to know itself?

Knowledge is the mission of universities, and yet most universities are so large and heterogeneous that they struggle to know their own research community. Given how many researchers there are at any university, and how independent they are, it is difficult to manage information about research activities, even though universities need current information to assess and promote programs, teaching and research. Knowledge about the interdisciplinary areas of research strength that cross departmental boundaries is even harder to find. What knowledge there is gets hoarded by units like departments that have to assess annual performance.

This paper describes the development of a research network map of the interests of the humanists, social scientists, and artists at the UofA (cloud.tapor.ca/viz/phil/), which is part of KIAS' project to understand where there were interdisciplinary strengths at the university and to help connect researchers. In this

[1] kias.ualberta.ca

paper we will demonstrate the Research Map, describe the challenges around gathering information at an institutional level, and close by discussing how it is now being adapted to be used by another interdisciplinary unit, the Digital Synergies research group.

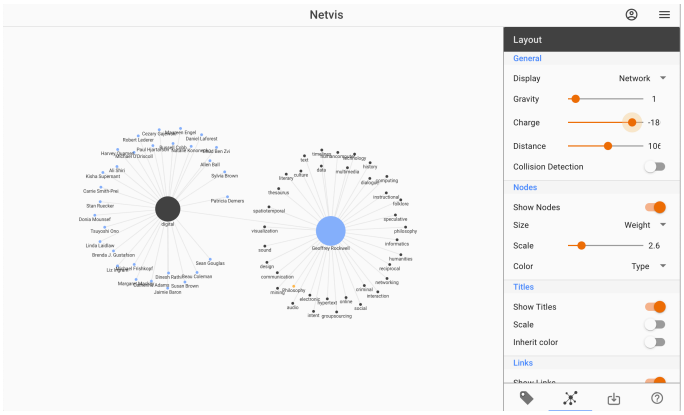


Figure 1: Research Map with "digital" and "Geoffrey Rockwell" expanded.

Research Map

The Research Map is a web-based network visualization interface that shows the connections and relationships between faculty members, their research interests, and their departmental affiliation. The tool allows for dataset exploration through straightforward searching, and a more serendipitous browsing by means of the selection of a few entities to be displayed, followed by an interactive expansion of its network (Fig. 1). The outcome is a visualization showing clusters of knowledge and webs of intersectionality, revealing not only the richness of the academic production but also the possibilities of future collaboration between scholars and departments.

The following snapshots describe the main Research Map features and illustrate some aspects of its interactivity and functionality. A video demonstration is available on Youtube: https://www.youtube.com/watch?v=zl6XCXmQfog&feature=youtu.be

Visualization

The visualization is interactive. You can move nodes around and unpack them. By default, faculty members are shown as blue circles, research interests as black circles, and the departments as orange circles. The size of each node corresponds to the number of relationships it has at the current state of the visualization. You can add more nodes by double-clicking on existing nodes or selecting tags in the sidebar. You can get more info about each node by clicking and holding on a node within the visualization, or on the little "i" icon on the tag list. The panel at the bottom shows the contextual data for each node and its relationships (Fig. 2).

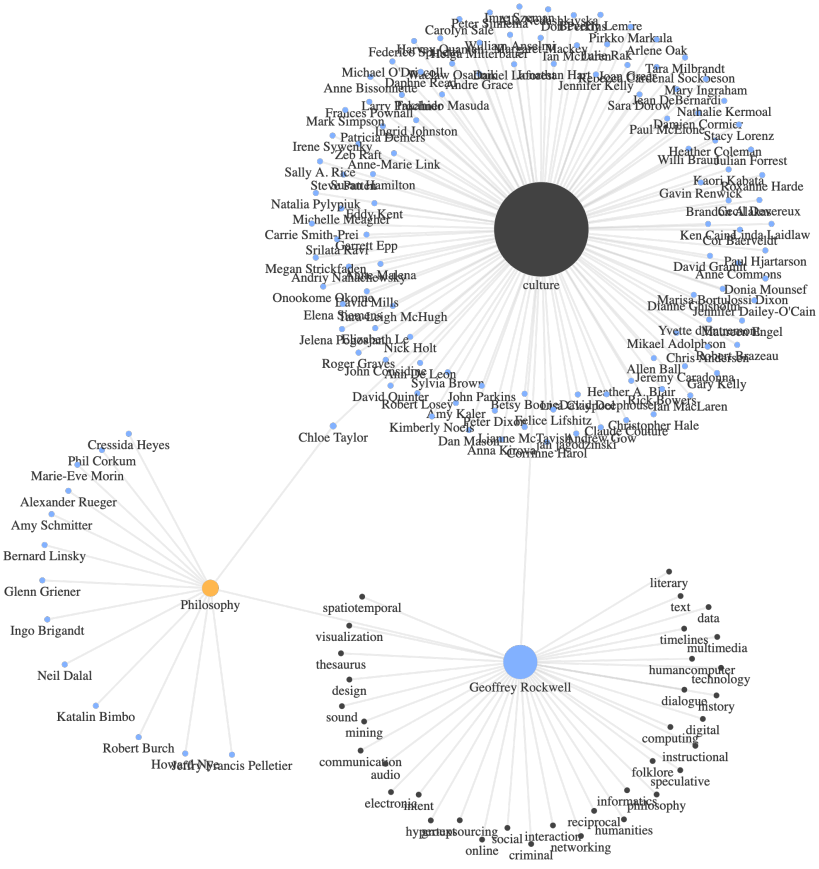


Figure 2: Research Map visualization.

Dataset

Collection of faculty members, research interests, and departments. The dataset is organized in alphabetical order and separated according to these categories. Use the auto-complete search box to find data quickly. Select as many as you want to add to the visualization (Fig. 3, on the left).

Layout

There are several options to control and customize the visualization: A. adjust the tension and distance between the nodes; B. Change the nodes' size, scale, and colour (by type or cluster); C. Choose how the nodes' titles will be rendered; D. Adjust links' thickness (using the number of connections between nodes), colour, and strength (Fig. 3, on the right).

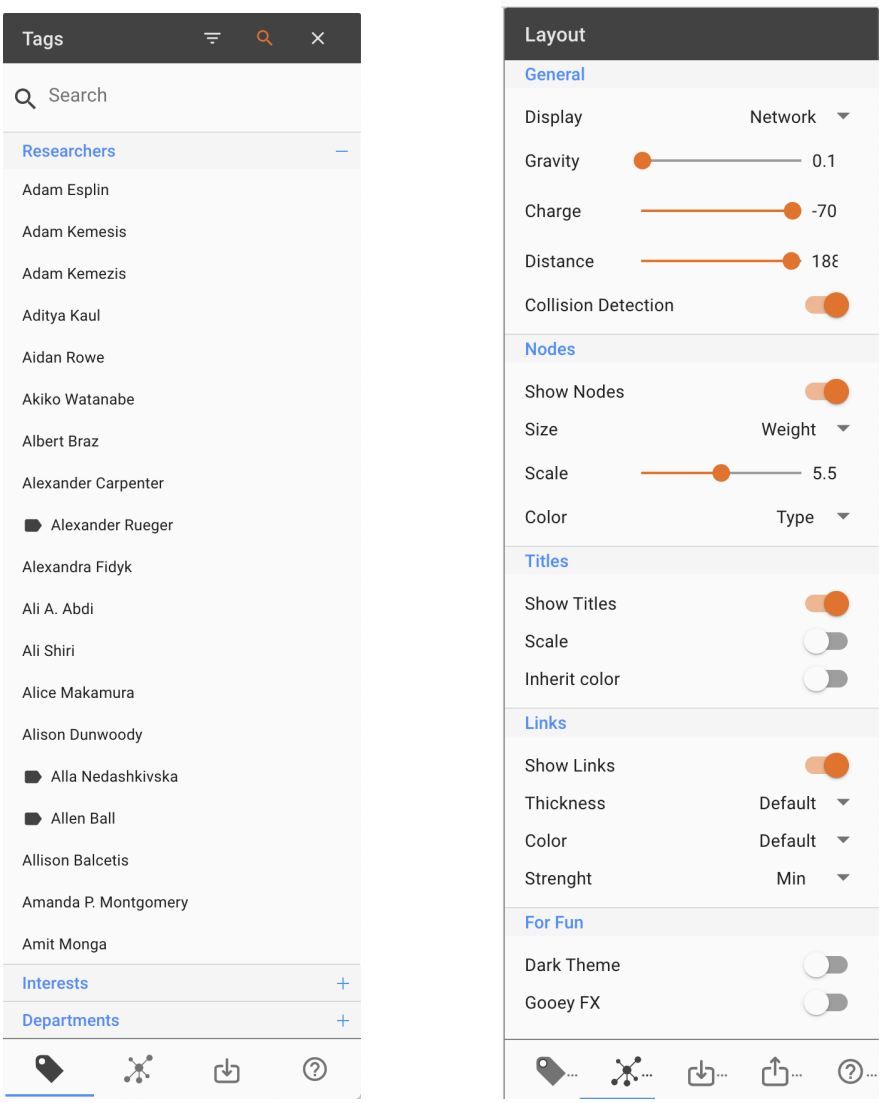


Figure 3: Left: Tags collection. Right: Layout customization

Data Management

It is possible to add new tags or edit existing ones: Select the type, give it a name, and add some relationships to other tags. You can also import a batch of tags, with support for CSV and JSON format (Fig. 4, on the left).

Export

The current state of the visualization can be exported as data (CSV or JSON format) and as an image (PNG or SVG format) (Fig. 4, on the right).

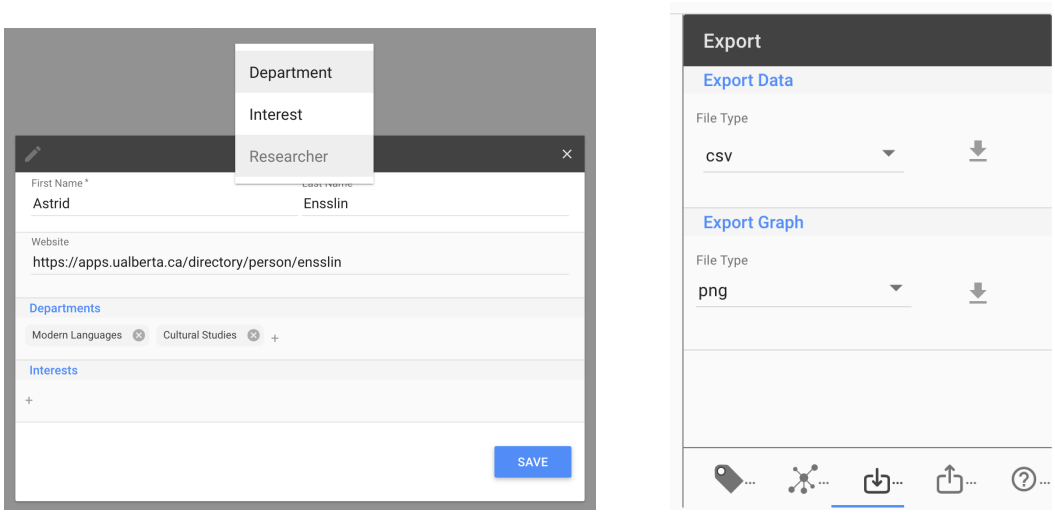


Figure 4: Left Add and edit data. Right: Export data.

Building a Research Map

Rationale

The genesis of this project was the need by KIAS to be able to represent social science, humanities, and arts (SSHA) research across campus. As KIAS was one of the few units that supported SSHA research across the university, the Director Geoffrey Rockwell was on University-level committees where he had to represent and promote the breadth of SSHA research and creative work without actually knowing what was happening or what we were good at, let alone being able to say what areas of interdisciplinary research were strengths. This is not surprising given that at a large university like the UofA there are about 700 SSHA researchers across at least 13 of the 18 faculties/campuses[2], each reporting their research to their own chairs and deans, but not sharing in any central forum. We literally did not know ourselves as an institution. What was worse is that some University leaders were pushing for the identification of specific areas of excellence to be formally identified and differentially supported, an exercise that most universities go through for strategic reasons. In response KIAS approached the Office of the Vice President Research (specifically the Associate VP Research in charge of the SSHA area) with a proposal to gather data and analyze it. The goals as described at the time included:

- To have a **sense of the breadth** of research across the university.
- To know **who does research** in specific areas when we are trying to connect researchers to each other or to the media.

[2] The Faculties that have SSHA researchers include: Arts; Education; Kinesiology, Sport & Recreation; Native Studies; ALES (Agriculture, Life & Environmental Sciences); Augustana campus; Campus Saint-Jean; Law; Business; Nursing; Rehabilitation Medicine; Public Health; and Extension.

- To know what **areas of strength** there are, especially those dispersed across faculties that do not seem central enough to any one faculty.
- To know who might be the **champions of research** areas willing to develop internal networks.
- To know what **areas of potential strength** are emerging so that we can provide support for the formation of research teams.[3]

There was a further tacit goal, and that was to try to show to senior administrators not from SSHA areas that there might be a very different distribution of excellence in the SSHA than in Science, Engineering and Medicine. Instead of having a small number of standout projects that were clearly recognized as "world class" by various forms of bibliometrics, in the SSHA one might find that excellence was a dense interdisciplinary network of overlapping projects around an area like Digital Humanities. We hoped that a research network map of SSHA might show the different configuration of fields and researchers in a way that encouraged understanding of the organic flourishing possible in the SSHA.

Process

To get to a research map we did the following:

1. **Scrape** the text of web pages of all full-time professors that were doing SSHA research.
2. To process the data to **generate research keywords** for each professor and for each department.
3. To generate **word clouds** for faculty and departments and to send the cloud images to the chairs of departments as a way of encouraging departments to help us correct our keyword data.
4. To generate a **large map** of all the keywords and their connections that could be distributed to research leaders as a poster.
5. To develop an **interactive online tool**, the Research Map proper, that could be used to explore SSHA research at the UofA.

A number of things can be said of this process. The first point to be made is that going through every relevant department in order to scrape the self-description of every full-time faculty member has to be one of the most inefficient ways to gather data about research at a university. All universities collect this data in some form of annual report for the purposes of assessment. The problem at the UofA is that the annual reports are treated as private despite being made up primarily of publications and other information about public activities. Other universities use internal reporting to automatically populate databases that allow research to be efficiently searched and summarized. Given the balkanization of systems at a university like the UofA most faculty who want a research presence on the web resort to services like Academia.edu.[4]

[3] This is from the proposal "What's Our Research? Research mapping initiative, v.4" prepared by KIAS in 2014. The emphasis (bolding) is in the original.

[4] academia.edu

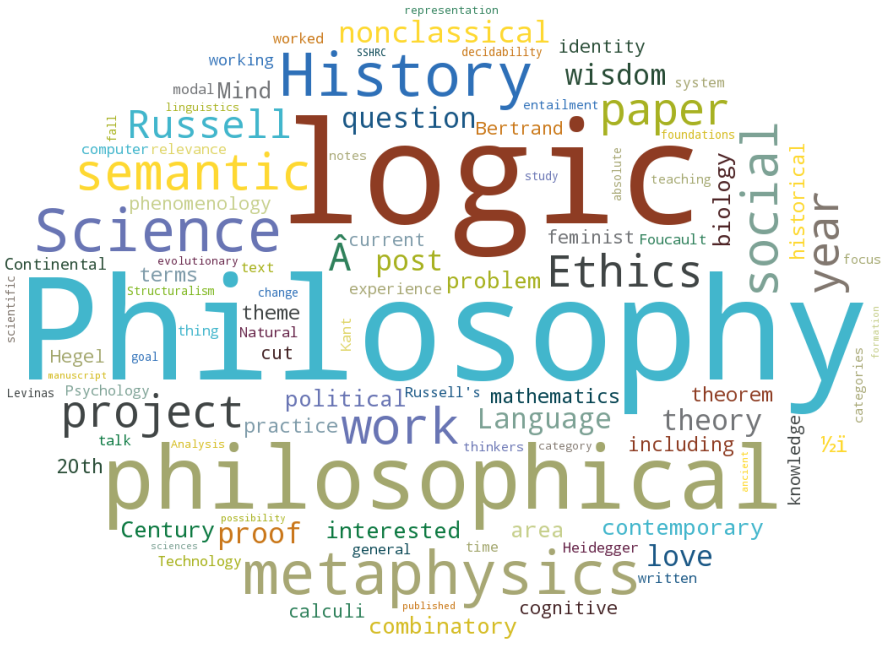


Figure 5: Word Cloud for Department of Philosophy.

Second, something that surprised us was the reception of the word clouds. Many who had not maintained their personal web sites were surprised by the way they were (or were not) represented online. For at least one department (Philosophy) our project was the incentive to encourage and support faculty to update their profiles which in many cases were dramatically out of date (Fig. 5). It should be noted that the problem of maintaining an accurate web presence is of growing importance to departments who want to recruit graduate students or be internationally recognized.[5]

The interactive Research Map was the last and most complex deliverable of this project. It was initially password protected so that research administrators could help us improve the data. Our original conception was inspired by the Explore Concordia site[6] which is an extraordinarily attractive and easy to use site that allows one to serendipitously explore topics and connections between researchers. We did not have the budget or remit to do something that sophisticated so we drew on the interactive design of a Media Studies PhD student, Luciano Frizzera, who had developed attractive prototypes when part of the INKE project and had a prototype for visualizing Twitter discourse around the Vancouver Transit Referendum,[7] which was similar to what KIAS wanted. Commissioning the Research Map from Frizzera had the advantage of supporting a student and getting an attractive interactive.

---

[5] See "Disciplinary Differences in Academic Web Presence – A Statistical Study of the UK", http://www.researchgate.net/publication/3211680l_Disciplinary_Differences_in_Academic_Web_Presence__A_Statistical_Study_of_the_UK

[6] explore.concordia.ca

[7] labs.fluxo.art.br/transit-debate

Maintenance

Like many digital humanities projects, the challenge of the Research Map lay not only in the development of the idea, but in the maintenance after the initial implementation. Maintaining the data has proven to be an annual updating project. Updating the data was one reason why KIAS teamed up with the Office of the VP Research. They have the authority to encourage people to send us updates and to disseminate news about the interactive. It was the Office of the VPR that distributed the word clouds to chairs and solicited feedback. This approach is predicated on the willingness and ability of the various stakeholders (VP Research office, departments, researchers) to collaborate, which may fluctuate over time as influenced by factors difficult to control (such as staff shortages, unforeseen crises, etc.).

Another avenue worth exploring for the on-going maintenance of the data is the scraping of academic social networks such as ResearchGate[8] or Academia.edu, already used by a large percentage of the targeted researchers — or in the future, of researcher-run alternatives to these for-profit sites such as ScholarlyHub.[9] Possible other sources of updated information consist of scraping and processing research data from the University's Research Archive (ERA)[10] and/or researchers profiles from ORCID,[11] but such approaches are still in an exploratory stage.

From a technical point of view, the actual process of updating the database has been significantly simplified. The most recent version of the Research Map allows administrators and editors to update the content of the database by uploading JSON or CSV data structured according to predefined specifications.[12] JSON is the preferred format as it can contain more information about each node. It also condenses all the data into a single file. The CSV file format was preserved as it is still somewhat more familiar and easy to grasp by updaters. The disadvantages of it are described below in the data model section. If CSV is used for database updates the information has to be broken down into two separate tables, one for nodes and one for their relationships. Once the updated info is loaded, the system will flag the new items in the tag lists.

Data Model

Part of the challenge of this project is to structure networked data to be stored and retrieved on demand. The dataset consists of three categories: Faculty Members, Research Interests, and Departments, with

[8] researchgate.net

[9] scholarlyhub.org

[10] era.library.ualberta.ca

[11] orcid.org

[12] The specification are available on this site: github.com/lucaju/netvis#add--import-data.

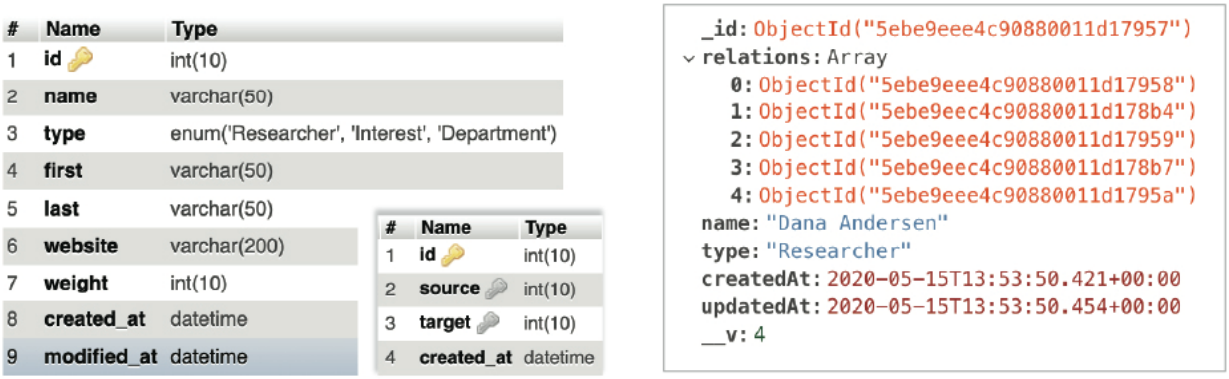
multiple relationships. These entries on each of these categories have just enough information to produce the visualization and provide context. Thus, Faculty Members are defined by their first and last name, their Departmental affiliations, a website, and a list of their research interests; Departments are defined by a name and a list of affiliated Faculty Members; and Research Interests are identified by a name and their connections to the Faculty Members.

This configuration reveals a clear hierarchical relationship proper of a university structure: Departments are composed of Faculty Members, and Faculty Members have Research Interests. There is no direct relation between Departments and Research Interests, except through their faculty members. That is, these are one-way relationships (source > target). However, this hierarchy is not exclusive since Faculty Members can be affiliated with more than one Department, and Research Interests can be connected to multiple Faculty Members from different Departments. Interestingly, this brings light to the interdisciplinary nature of the SSHA fields. Following these apparent contradictory patterns on the dataset, we were faced with a major question: how to structure the database to allow multiple connections between hierarchical entries?

We started the first version of the Research Map storing the dataset in two CSV files. It quickly became clear that this approach was not fitted and scalable for the job, and we need database storage. Our first choice was to use a relational database (MySQL), which provides a store of related data tables. We decided to store the entries in a single table with an attribute to identify the category: every row stores a different record (Fig. 6, on the left). This choice was thought to be a good fit for its simplicity (a single table for multiple categories) and scalability (in volume). A second table was used to store the relationship between the entries using their Unique Identifier (ID) as pairs (Source > Target).

However, because SQL debases design is rigidly tied to a data model, and the different categories have different attributes, the single table approach can be inadequate and inefficient. For instance, because Faculty Members might have a website attached to their record, all Research Interests also need to hold some value (*e.g.*, NULL) for this same attribute. This issue tends to worsen as the categories' complexity increases, such as when a faculty member has more than one website, or the possibility to the add more exclusive attributes to one of the categories.

An easy way to solve this problem would be to create a separate table for each category. However, this would produce another scalability issue: for each additional category we decided to support in the future, universities, research centres, and funding agencies, for example, new tables would need to be designed (with further production of code to handle them). Thus, the question becomes how to allow flexibility in each category's complexity and, at the same time, keep the scalability in volume?



#	Name	Type
1	id	int(10)
2	name	varchar(50)
3	type	enum('Researcher', 'Interest', 'Department')
4	first	varchar(50)
5	last	varchar(50)
6	website	varchar(200)
7	weight	int(10)
8	created_at	datetime
9	modified_at	datetime

#	Name	Type
1	id	int(10)
2	source	int(10)
3	target	int(10)
4	created_at	datetime

Figure 6: Left: Data structure in a SQL relational database. Right: Data object in a NOSQL database.

We are experimenting with NoSQL databases (MongoDB) for the second version of the Research Map (currently in development). On MongoDB, similar documents can be stored in a collection, which is analogous to an SQL table. However, there is no rigid schema (predefined columns) to be followed; collections can store any data you like in any document. We take advantage of this anti-pattern paradigm to create a single collection to hold all the entries with just the shared field tagged as required: name, type, relations, and timestamps (Fig. 6, on the right). Extra fields can be optionally added for each entry.

With this configuration, Research Interests do not need to have NULL as a website attribute; Faculty Members can have multiple websites and extra metadata; new categories of entries can be incorporated without the need to create new collections (tables) or required fields (columns). The relationship between the entries is stored within the entries themselves, through a reference to their ID. The downside of this approach is that it requires that both ends of the relationship store the information; otherwise, just one of them (the source) would "know" about the other (the target). However, this could be solved by introducing a new collection to hold these relationships, similar to the second table in the SQL relational database.

Generalizing

The first version of the Research Map was strictly set up to store, retrieve, and visualize a specific set of categories: Departments, Faculty Members, and Research Interests. We quickly identify the potential of such a tool to be used by others, in different contexts, and even with a wider variety of categories. In the effort to generalize the Research Map, we first introduce the capability to "import" datasets. With the data appropriately formatted, the tool could be used by other research groups or universities. To facilitate the adoption, we develop an easy-to-setup process with minimum technical skills required.

The change in the data storage approach discussed above constitutes a further step to generalize the Research Map. The next step should be a redesign of the user interface to enable all the possibilities

triggered by the new data structure. This last step will make it possible to extrapolate the initial purpose of this tool and expand to other types of relationships (*e.g.*, scholarly citations, research collaboration, institutional ties, social-network relationships, text-analysis).

Adapting the Map

We are now in the process of adapting the Research Map to be used by others within the University like the Digital Synergies research group.[13] It is a "signature area" for research and creative collaboration focused on digital society, digital methodology, and digital literacies, bringing together researchers from a variety of disciplines. Adapting the Research Map to be embedded in a website streamlines the process of organizing and visualizing the connections between researchers. We are, in effect, using digital social network analysis methods to help people understand the interdisciplinary network itself.

Our process of adapting the Research Map began with collecting the data of researchers within the Digital Synergies sphere and adapting the code to fit our needs. Joining the Research Map is completely voluntary and we collected data such as department, research interests which are based on the SSHRC key word ontology, and website links to any relevant projects. This data is processed into a spreadsheet that indicates the nodes and edges of the Research Map.

Our Digital Synergies Research Map[14] is linked through our website and displays an instance of the most recently updated Map. Users can navigate the map by choosing the researcher's name, a department, or a research interest. All related information will show up for the selected criteria. Users who wish to be a Digital Synergies researcher featured on our Researcher Map can register to do so through our website,[15] which is updated and maintained regularly to the most recent information. Moving forward, we will update the Map to show information such as if the researcher is looking for graduate students, if they are looking to collaborate with other researchers on projects, and their primary area of research.

Concluding remarks

In times of austerity and continual budget cuts, intra-organizational knowledge and communication are more critical than ever for maintaining a healthy, creative, and mutually beneficial research landscape. This paper has showcased a tool designed to help researchers and university administrators identify previously undetected and potential links and synergies, especially for projects that require a strong interdisciplinary component. The tool may come in useful for future institutional exercises in signature area creation and development, not least because it enables both bottom-up and top-down approaches. A

[13] digisyn.arts.ualberta.ca/home/main

[14] digisyn.arts.ualberta.ca/home/researcher-map

[15] digisyn.arts.ualberta.ca/home/new-member-form

future step in the development of the UofA Research Maps would be to link them to more information about faculty, following a consistent, coherent visual design agenda, similar to Explore Concordia, for example, and to integrate them into a more rigorously coordinated network of research information management tools. Some institutions are using tools like Duraspace Vivo[16] to create a common integrated scholarly presence.

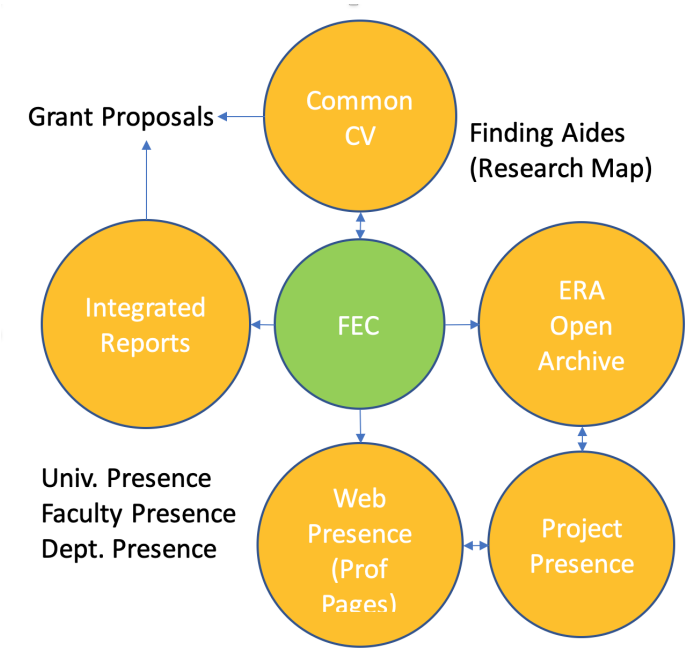


Figure 7: Model for Coordinated Research Presence.

One of the lessons learned trying to gather the data for the Research Map is the lack of coordination between the different data silos at the UofA (Fig. 7). The annual Faculty Evaluation Committee (FEC) reports are in their own silo which means that faculty need to reenter the same data in their own Professor web pages, in web sites like Academia.edu, and in the standard Canadian CV forms like the Common CV which is used for grant applications. Nor is there any link to the Library's service that openly archives faculty research. Departments and individuals have to manually archive their publications. If you are in a research project with its own web site, like Digital Synergies, that is yet another place you have to update your information. This situation is partly due to the independent evolution of the different systems on campus, but also our complacency. Everyone is content to promote their research through their own venues and few see the need to have a coordinated system.

Needless to say, we can imagine a more coordinated system as mapped out above or like Vivo mentioned above. Such coordination is not a dream; other jurisdictions have implemented large-scale coordinated

[16] duraspace.org/vivo/

systems like the Research Map[17] run by the Japanese Science and Technology Agency (JST) that was designed by the National Institute for Informatics (NII). This service "comprehensively collects and provides data on research institutions and researchers, etc. relating to the universities and public institutions in Japan."[18] Such large-scale systems ideally would have an open Application Programming Interface (API) that would allow experiments like the Research Map to build on top of well-maintained institutional research data.

References

Moran, J. (2010). *Interdisciplinarity*. London, Routledge. 2nd Edition on Kindle.

Thelwall, Mike & Price, Liz. (2013). Disciplinary Differences in Academic Web Presence – A Statistical Study of the UK. Libri. 10.1515/LIBR.2003.242.

[17] researchmap.jp

[18] https://researchmap.jp/public/about/operations