# New Mexico Data Sharing Project






## FINAL REPORT LSC TIG #15062

# Table of Contents

LEGAL SERVICES CORPORATION
TIG FINAL REPORT

Grantee Name: New Mexico Legal Aid
TIG Grant #: 15062
Submission Date: April 30, 2018
Contact Person: Ed Marks, NMLA Executive Director
Telephone: 505.768.6122
edm@nmlegalaid.org

## I. Project Goals and Objectives

New Mexico Legal Aid's Data Sharing Project, first launched in 2012 with support from TIG #12018, has been a key step in furthering NMLA's commitment to using data-based evidence to better understand client needs, to more timely identify emerging issues, to proactively address individual factors and systemic patterns that further complicate client legal issues, and to encourage stronger multi-agency collaboration within New Mexico's legal services provider community. Through this second phase of the project supported by TIG #15062, we worked with data analysis experts to develop sophisticated data analysis tools that provide a deeper understanding of data relationships and tests of statistically significant changes within aggregate databases to ferret out patterns and causal relationships that would otherwise likely remain unrecognized through traditional data reporting methods.

The primary goal stated for the project was to "Incorporate knowledge management tools that will empower legal services organizations to understand and more proactively address correlations between clients and legal issues and to more effectively prioritize and execute responses and strategies." The specific objectives were as follows:

- Continue multi-organization partnership, collaboration, and data sharing to ensure continued issue-oriented collaboration and coordination on strategic advocacy efforts based on shared data.
- Implement four previously identified improvements to existing data sharing system that will strengthen the system and make data analyses more robust.
- Identify data questions, analysis methods, and relevant data sets (including external data) to support robust analyses that may be used by legal services organizations to more effectively and proactively serve client communities by identifying correlations with client or community characteristics in relation to emerging trends or clusters occurring within common legal problems.
- Identify and implement data analysis software that is the best fit in terms of cost, ease of use, and applicability for NMLA and other legal services organizations that might want to follow our lead.
- Use the new tool to foster improved responses for individual clients, groups of clients, and the entire client community, and to foster issue-oriented organization-specific strategies, and issue-oriented collaboration and coordinated strategic advocacy among partner organizations.
- Create deliverables that inform other legal aid organizations about our multi-organization collaboration, partnership with data analysis experts, analysis strategies, knowledge management tools, external data, and practical application of the new tool

The project goals and objectives were successfully achieved.

## II. Evaluation Data and Methodologies

**Evidence of Multi-Agency Partnership:** We continued our collaborative partnership with New Mexico-based civil legal services provider agencies which also had participated in the initial TIG 12018 phase of the project: Law Access New Mexico, Pegasus Legal Services for Children, the State Bar of New Mexico's Legal Resources for the Elderly project, and the Senior Citizens Law Office. We held group meetings with these partner agencies throughout the project timeline, including meetings in which we tested several iterations of the analyses reports and documented partner feedback.

**Evidence of Improvements to Data Sharing Site:** We completed more than the original four planned improvements, including an updated Trends tab, a new Problem Codes Trends tab, and an improved method for downloading the entire database. We listed the improvements in meeting minutes, and included new fields in our analyses reports. After sending 41 automatic email alerts between July 2016 and May 2017, we asked for feedback on the utility of such real-time alerts. Based on their responses we ended the automatic alerts and shift to quarterly distribution of the new data analyses reports. Project partners still may access the Data Trends site as often as they like for target review of specific issues.

**Documentation of Data Issues, Data Questions, Analysis Methods, and Data Sets:** We spent considerable time working with our partner organizations to ensure that all the internal data included in our shared data system was mapped definitionally. Because of customized variations in data fields that some partners had added to their systems, the solutions to the mapping issues required complex coding both within the data sharing site and in the analyses reports themselves. All coding was documented. With assistance from our data experts, we documented data questions, analyses methods, internal and external data, and methods for linking and updating data in instructions and meeting minutes.

**Documentation of Tested Data Analysis Software**: We compared various open-source and proprietary data analysis software options by cost, learning curve, and various other factors and documented our findings. Based on Pika software restrictions, cost and learning curve issues, we decided that the best solution was to use Crystal Reports, which NMLA was already using. Our consultant previously had created a Crystal Reports training guide specific to NMLA, and has trained NMLA staff on the software.

**Documentation of Analyses Results:** We were delayed in starting our analyses by the time-consuming task of definitionally mapping all fields across participating organizations. These data issues placed some limits on the types of analyses we could use. Still, we could apply the Two Population Difference of Proportion Test to test for statistically significant changes in selected data over a specified timeline. These steps were documented in the minutes of project meetings. The tools and reports developed by this project can now be rolled out to all staff to improve the impact of our work. But the extra time required to develop these tools limited survey responses to selected agency staff.

**Deliverables Presented:** We presented guidelines about our multi-agency collaboration, partnerships with data analysis experts, data questions, methods of analysis, data sets, definitional data mapping issues, strategies for accessing external data via an API, analysis software employed, aggregated data analyses report results and practical applications at LSC's ITC Conference held in New Orleans in January 2018. We fielded multiple questions during and after the presentation from attendees that indicated interest in replicating both the collaborative partnerships and types of analyses we shared.

## III. Summary of Major Accomplishments, Recommendations and Future Steps

New Mexico Legal Aid (NMLA) partnered with Law Access New Mexico, Pegasus Legal Services for Children, the State Bar of New Mexico, Senior Citizens Law Office, along with Strategic Data Analytics and Cleveland State University to create reports that aggregate and analyze internal data from the five legal service organizations and external U.S. Census data about the eligible client population in New Mexico. The reports provide a wealth of information about concurrent changes across legal problems, across various client groups, and across eligible clients within different geographic regions.

The most significant factors influencing success of this project were our partnerships with data analysis experts as well as with allied legal services organizations. The data analysis experts helped us effectively organize and leverage mountains of internal and external data, address the variations in field definitions by definitionally mapping fields, identify relevant data questions, formulate fitting analyses, identify the best cost-effective software, and communicate our findings in user-friendly formats. Our legal services partners agreed to share their non-confidential client and case data and acted as ongoing beta testers as we formulated our analysis strategies and developed user-friendly and informative reports.

As previously indicated, this project grew out of a previous TIG project (TIG 12018) during which the five participating legal service organizations developed a secure web-based data sharing system to analyze aggregate non-confidential intake data. That initial iteration of the data-sharing system offered only limited options for analysis strategies. We knew more could be done with the available aggregate data.

Before launching into new forms of analyses, we first had to resolve many data variations and data field definitional nuances specific to each individual organization. With input from our partners and multiple rounds of testing, we implemented complex data definitional mapping code that ensures our reports are meaningful and provide useful “apples to apples” results to all our partners. Because of our experiences, we recommend that any project aggregating data across multiple organizations allot sufficient time early in their project to ensure data fields are adequately defined and mapped.

Although the data variations across the partner organizations limited our ability to confirm formal correlations across fields, we nonetheless could apply sophisticated statistical analyses to better understand when changes over time among various client and case characteristics were statistically significant and warranted further investigation. The project’s partnership with data analysis experts helped identify the best and most informative statistical test available to understand our data with all its nuances. This turned out to be the Two Population Difference of Proportion test, a method that would be unknown to most legal services staff without the ability to partner with qualified data experts. We recommend that any future data analysis projects partner with data analysis experts knowledgeable about multiple methods of statistical analyses as early as possible in the project timeline, so that regardless of data issues the most informative and applicable analyses methods will be applied.

We are excited to continue using our internal data Problem Code Profile Reports and our external County and Zip Code Data Analysis reports to better understand our clients and meet their needs. We look forward to continued collaboration with our partner agencies and the addition of more external datasets to our system and analyses reports.

# IV. In-Depth Analysis of Accomplishments

## A. Project Goal

The original New Mexico Data Sharing Project (TIG #12018, completed in 2014) developed a secure web-based data sharing system to analyze non-confidential intake data daily. The system generated charts to indicate emerging issues and trends that previously would have been overlooked other than through anecdotal discussions or case by case review. Still, the system was limited in that it identified trends and patterns occurring within single issue areas. Without some significant degree of manual filtering of data and individual interpretation of the resulting data sets, the system was not able to provide accessible information about changes across multiple legal problem or client demographic fields. Nor was it able to provide any indication of statistically significant data relationships or changes.

As previously described, the current TIG project (15062) addressed these limits by developing expanded analyses tools focused on identifying concurrent changes across legal problems, across various client groups, and across eligible clients within different geographic regions.

## B. Project Objective 1: Multi-Organization Partnership

Our first project objective was to continue our ongoing multi-organization partnership, collaboration, and data sharing to support continued issue-oriented collaboration and coordination on strategic advocacy efforts based on shared data.

### 1. Partner Participation

We held a series of meetings and conducted regular communication with partner organizations to support continuation of their active participation in the project. To start, we re-convened the partner group originally formed for the predecessor TIG 12018 project, including Ed Marks, New Mexico Legal Aid; Carol Garner and Conrad Rocha, Law Access New Mexico; Liz McGrath, Pegasus Legal Services for Children; Stormy Ralstin and Maria Tanner, State Bar of New Mexico; and Ellen Leitzer and Kathy Heyman, Senior Citizens Law Office. (Note that the Southwest Women's Law Center decided not to actively participate in the project due to its limited client caseload and focus on a small range of specialized legal issues. Also, note that Bette Fleishman became the new Executive Director and project contact at Pegasus in September of 2017).

We held a kick-off meeting with our partner organizations in March of 2016 to which representatives of the following potential future partner organizations were also invited to attend: Enlace Comunitario, NM Immigrant Law Center New Mexico Center on Law and Poverty, Southwest Women's Law Center, DNA People's Legal Services, New Mexico Immigrant Law Center, Disability Rights New Mexico, Native American Disability Law Center, United South Broadway Corporation, and the University of New Mexico School of Law. At this meeting, all attendees shared data questions, the answers to which they believed could help them improve client service. Those who had seen the data sharing website developed under the predecessor TIG project also shared suggestions for improvements to the website.

Through a series of follow-up meetings, shared Google documents, and one-on-one discussions, our project partners provided valuable input regarding their goals for using data more strategically, their insight about the needs of eligible clients in New Mexico, and ideas for additional improvements to the data sharing website. The project meetings and follow-up communications regarding resolution of variations in data field definitions among the partner organizations (over 100 emails and phone calls

between June 2016 and November 2017) supported our work to write code that successfully mapped fields across organizations.

Additional input and communications from our partner organizations led to several requests and recommendations, including: there should be a threshold number of intakes within each legal issue problem code so that data will not be automatically flagged for review of numerical shifts within particular low-volume legal problem codes; filters in the system need to display plain language field names in addition to numerical codes (e.g., 01-Bankruptcy, rather than just 01); the trends and reports tabs needed more explanatory headings and explanations of what data are included; and the system should provide analyses for: a) all New Mexico counties, b) Bernalillo County (Albuquerque metro area) alone, and c) all counties excluding Bernalillo so that the high population density in the Albuquerque metro area does not unduly overshadow data from less-densely populated rural areas of the state; c) provide analyses for each partner agency individually in addition to the multi-agency aggregated versions.

## C. Project Objective 2: Implement Data Sharing System Improvements

To ensure high quality analysis upon which our partners could rely, we spent significant time identifying and resolving issues to prioritize needed improvements to the data sharing system. Such potential improvements that were identified included addition of new fields and new filters to the system; the ability to set up automatic email alerts; and development of methodologies to help partner organizations compare aggregated system data to their own internal confidential data. Based on extensive review, we identified and implemented those and additional improvements, including updating the Trends tab, creating a new Problem Codes Trends tab, and incorporating a method for downloading the entire database.

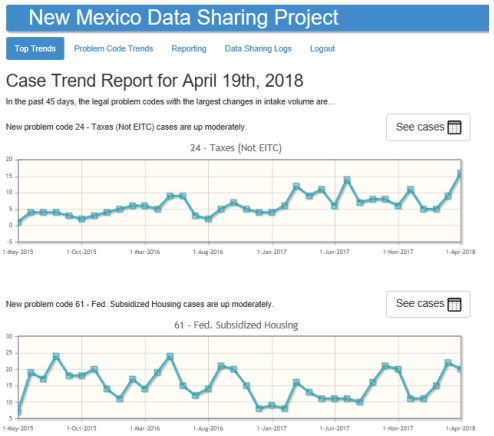
### 1. Work with SDA and Pika to Implement Improvements

Rachel Perry from Strategic Data Analytics (SDA) spent significant time acquiring input from partner organizations and worked with Pika to ensure the implemented changes met the needs of the partners. Implemented improvements to the data sharing system include:

a. **New Fields:** We identified new fields to add to the system and collected data definitions from each organization so that we could map the fields correctly. The new fields included: number of children; number of persons helped; poverty level; language spoken; veteran in household; age at intake; case closing code; and numerical client_id. Other fields that were considered, but not included were: educational level; type of income; citizenship status; domestic violence victim status; reason case was rejected, and female head of household. These fields were rejected either because they are not collected by all the partner organizations or because of confidentiality concerns that some fields could unintentionally lead to identification of client identities or other unintentional disclosure of confidential information.

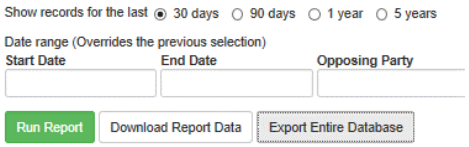
b. **New Filters:** From among the new fields, filters were added for fields with limited discrete options: veteran in household, language, and outcome. Because we added the closed code field late in the project, a filter was not added even though this is also a field with limited discrete options. Of course, close code, along with all the other new and existing fields, appears in data exported from the system and thus, filtering outside of the system using this or any other field is possible.


c. **Updated Trends Tab:** The Trends tab was renamed to Top Trends and includes graphs for the problem codes with the top ten highest percentage change over the last three years in descending order. Additionally, we applied a threshold so that graphs will only appear if a problem code has at least ten cases in the most recent period. Labeling on the page was included to clarify the contents of the page. Finally, we removed the graph for cases with blank problem code. Results of Trends Tab were tested against raw data.

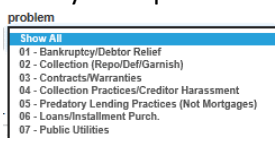



d. **Created New Problem Code Trends Tab:** This new tab shows trends for all problem codes and no minimum threshold is applied so staff who are curious about trends for particular problem codes can find all of them on this tab. The graphs are sorted in problem code order so that particular problem codes are easy to find. Cases with blank problem codes are included on this tab so that partner organizations have access to data about the volume of cases missing problem codes. Labeling on the page was included to clarify the contents of the page. Results of Problem Codes Trends Tab were tested against raw data.

e. **Improved Access to Data with New Export Entire Database Button**: This new button on the Reporting tab allows for a CSV download of the entire database. Previously when reports were run, users could only see a maximum of 10,000 records on the screen and only download that many. With new access to all the records, more complex analyses are possible. Data were downloaded and tested against program data from NMLA.




f. **Improved Data Labeling on Reporting Tab:** We implemented new coding to exclude erroneous or very infrequent data (options with fewer than 10 entries do not show up), which has significantly improved the data options for the Gender, Zip, and County fields. We also improved labels that used to show data options (such as 1, 2, 3, etc.) so that they now show substantive options (such as 01-Bankruptcy/Debtor Relief, 02-Collections, 03-Contracts/Warranties, etc.). These improvements were applied to the following fields: Race, Hispanic, Disabled, Age Over 60, Problem, Outcome, Veteran in Household, and Language.




g. **Automatic Email Alerts:** An email alert was set up to go out once a week and show the top 15 problem codes in terms of change in volume over last three years. Between July 2016 and May

2017, 41 email alerts were distributed. After receiving these email alerts, partners reported that the they were too frequent and not informative enough. In addition, our new analyses reports are significantly more useful, providing detailed information about legal problem codes with statistically significant demographic changes. Thus, we decided to distribute the results of our new analyses to each of the partner organizations once per quarter. Project partners are still free, however, to access data updates as often as they wish.

h. **Confidential Data Fields:** We discussed having organizations upload confidential data to the system that they could then use along with the non-confidential aggregated data for analysis. Significant concerns arose about confidentiality and complex programming required in Pika. Thus, we decided against this idea and instead created organization-specific analyses reports that will be distributed once per quarter along with the aggregated, system-wide reports. Also, each organization can download aggregated data from the system at any time using our new Export Entire Database button and then compare it to confidential internal data.

## D. Project Objective 3: Identifying Data Questions, Analysis Methods, & Data Sets to Support Robust Analyses

### 1. Work with Experts

Rachel Perry, who has been providing data analysis services to NMLA for several years, was a key leader of this project. Ms. Perry has over 20 years of experience conducting data analyses, eight of which have been on behalf of legal aid organizations. Having previously worked at the Legal Aid Society of Cleveland, she understands legal aid data structures, data strengths, and data challenges and has a deep knowledge of our case management systems, Pika in particular. Her understanding of the data needs of legal aid organizations proved to be invaluable to this project as she worked seamlessly with all partner organizations, served as liaison between the partners and Pika, designed and conducted all data analyses, prepared all data reports, and shared the project findings with the larger legal aid community at the most recent LSC-sponsored ITC conference in January 2018. Additionally, Ms. Perry has strong connections to academic researchers and thus was able to help us find an important academic data analysis partner.

A review of potential partners from the University of New Mexico indicated a disproportionate focus on criminal justice issues in the relevant departments rather than the civil legal issues that were the focus of this project. Additionally, many of the centers at UNM require payments for service. Therefore, Ms. Perry reached out to Professor Brian Mikelbank from Cleveland State University's Maxine Goodman Levin School of Urban Affairs. Professor Mikelbank is familiar with legal services data, having previously worked on a TIG project for the Legal Aid Society of Cleveland and Montana Legal Services Association. Professor Mikelbank met with Ms. Perry on multiple occasions and provided significant assistance identifying the most appropriate type of statistical analysis based on our data constraints, helped to design the analyses, and helped to interpret our results. He graciously provided all this assistance as a volunteer.

### 2. Developed Data Questions

Throughout the project, we were focused on an overarching goal to uncover patterns in our aggregate data that might indicate links among and between clients' legal problems and demographic characteristics. The first iteration of the data sharing system provided us with trend data limited to increases and decreases in intake volume by problem code, but it did not indicate whether those spikes or dips were statistically significant. In addition, there was no threshold in place, meaning that

frequently we were seeing trend graphs that might cause alarm because of a seemingly enormous proportional change, but were based on a change of only one or two cases for a legal problem with very few total cases over time. Our main data question became: Which changes deserve attention, i.e., which changes are statistically significant? A key set of sub-questions then became: What else is happening in terms of indications of data relationships that correlate or suggest causation with multiple legal problems, client demographics, and eligible community demographics?

## 3. Identified Analysis Methodologies

With assistance from Professor Mikelbank and Ms. Perry, we reviewed multiple types of potential analyses relevant to categorical and ordinal data, including Chi-Square Goodness of Fit Test Frequencies, Chi-Square Contingency Table Analysis, Spearman's Rank Order Correlation, and Multinomial logistic regression. For several reasons, these more sophisticated forms of statistical analyses would not produce the desired results, in part because not all participating organizations collected the same sets of data fields. For example, some data results were skewed because two of the partner organizations focus only on seniors and one organization focuses only on children. Some data also are in check box format in their organizations' databases making it unclear whether the case handler intended that an unchecked box means "no" or "blank." The various organizations also have different instructions and/or different interpretations of certain data fields. These types of variations could cause the analyses results to report erroneous data relationships because of the skew present in the data.

Professor Mikelbank and Ms. Perry determined that, based on the limitations listed above, it would be best to apply the Two Population Difference of Proportion test to compare the share of clients that correlate with particular legal problems from different demographic groups in the last 90 days compared to the last three years. They further determined that the results of this analysis would be most effectively shared with the organizations via a series of legal problem code profiles.

The first iterations of the Problem Code Profile Reports included Two Population Difference of Proportion tests applied to each binomial field (Ethnicity, 60 & Older, Disabled, Children in Household, Veteran in Household) and simple Frequency Comparisons for the categorical fields (Race, Zip Code, Age Range, Gender, Language, Poverty). In the final iteration of the profile reports, we were able to apply the Two Population Difference of Proportion tests to all fields, comparing counts for the binomial fields and percentages for the categorical fields.

Our Two Population Difference of Proportion Tests included the following steps:

1. Data Preparation: Before implementing the tests, we excluded blank data from all fields and created categories for Age Range and Poverty Range.
2. Null Hypotheses: There is no difference in the 90-day proportion and the 3-year proportion of clients who (are Hispanic, are 60 & Older, are Disabled, etc. or who have the proportionally highest Race, have the proportionally highest Gender, have the proportionally highest Poverty Range, etc.) for each legal problem.
    a. Note that the 3-year period actually includes data for 33 months because data from the 90-day period cannot be counted in both populations. For convenience, we refer to the 33-month period as the 3-year period.
3. Formula 1: *Population Estimate:* $P_u=(N_1P_{s1}+N_2P_{s2})/(N_1+N_2)$
4. Formula 2: *Standard Error:* $\sigma_{p\text{-}p}=\sqrt{P_u(1-P_u)} * \sqrt{(N_1+N_2)/(N_1*N_2)}$
5. Formula 3: *Z-score:* $P_{s1}+P_{s2}/\sigma_{p\text{-}p}$
    a. If the Z-Score is between -1.96 and +1.96, the Null Hypothesis is True. There is no difference in proportions, or any difference that occurs is due to random fluctuations.

b. If the Z-Score is <-1.96 or >1.96, then the Null Hypothesis is False and there is a statistically significant difference in the proportions. The difference is not random and it's worth investigating.

6. If the Null Hypothesis is False, cross tabs appear in the report comparing the field with a statistically significant recent change to other demographic fields.
    a. For example, if the Disabled field shows a statistically significant change, a page of cross tabs appears showing the proportion of Disabled clients by Race, by Hispanic indicator, by Age, by Gender, by Children in Household indicator, by Language, by Veteran in Household indicator, and by Poverty Range.

## 4. Identified Internal Data & External Data

### a. Internal Data

Our starting point for internal data was the data already being uploaded to the system: gender, race, Hispanic, disabled, age over 60, zip, county, case_id, open date, close date, problem, opposing party, court name, judge name, and outcome. We reviewed each field and uncovered challenges with many of them that made analysis very difficult, and in some cases impossible, such as outcome data that ranged from 1 to 14 and from A to W with no way of knowing what those numbers or letters indicated. We knew we would have to tackle those issues, but we also knew we wanted to bring in additional data, so we decided to first select additional fields and second to work on getting both the existing and new fields into formats that would allow for meaningful analyses.

Each partner organization has its own mountain of internal data. Early on we also discussed possible new fields for inclusion in the system. We hoped to add fields that would give us deeper insight into our clients and their needs. The fields we considered were: age at intake, annual income, type of income, citizenship status, client_id, community partner, domestic violence involved, educational level, English fluency, food stamps, frail/needy, HIV/AIDS, household composition, intake type, language, marital status, number of children in household, number of persons helped, reason rejected, referral source, rural, special problem code, total debt, type of residence, woman as head of household, and youth in foster care. While all these fields would have been enormously informative, we could not add all of them because they are not collected by all of the partners and because of significant variation in how some of the fields are defined within the partner organizations. We settled on the following fields to add: number of children, number of persons helped, poverty level, language, veteran in household, age at intake, closed code, and client_id.

We then focused on issues around how the fields were brought into the system and how to map them definitionally so that we could perform meaningful analyses. While we knew there would be some field mapping issues, we did not foresee the extensive nature of the field mismatch across organizations, nor the significant amount of time it would require to write complex coding and formulas to ensure consistent categorization of and apples-to-apples comparisons of intra-organization data.

In particular, race, ethnicity, legal problems, outcomes, language, and close codes required complex definitional mapping. Additionally, some fields that allow users to enter textual answers are riddled with spelling issues or issues around how data are identified. For example, the County field has many misspelled entries, making it difficult to compare county-specific data. Another example is the Court field: some entries are spelled out (Fifth Judicial District Court), some use numbers (5th Judicial District Court), or abbreviations (5th Jud. Dist. Ct.). Formulas were written to address these issues.

One of the most challenging field mapping issues arose from an organization (SCLO) that uses two different sets of legal problem codes. One set is almost identical to the LSC problem codes, but the other (which they use for most of their cases) is quite different. We wrote and tested programming to map problem codes by their meaning within the data sharing system, and after months of trying to get this to work within the data sharing system, determined that we had to upload both sets of problem codes and implement the coding to map the fields in Crystal Reports, rather than within the data sharing system. An example of the legal problem coding in place within Crystal Reports to aggregate cases is for Medicaid cases: If LSC Problem Code = 51-Medicaid OR SCLO City Problem Code is one of [25-Medicaid, 28-Medicaid Waivers, 29-Institutional Medicaid Eligibility], then use problem code label "Medicaid").

This field mapping process required many rounds of testing to ensure that it is working. In fact, we provided the first field mapping definitions to Pika in June of 2016 and had to download and test the data, run results by our partner organizations, work with Pika on edits and corrections, and develop mapping formulas in Crystal Reports through September of 2017.

**b. External Data**

We reviewed multiple potential sources of external data, including the University of New Mexico Institute for Social Research, the University of New Mexico Statistical Analysis Center, the University of New Mexico Resource GIS Program, the University of New Mexico Bureau of Business & Economic Research, the University of New Mexico Center for Applied Research and Analysis, the City of Albuquerque Open Data, New Mexico Voices for Children, New Mexico Community Data Collaborative, and the State of New Mexico's Indicator-Based Information System (NM-IBIS). Many of these sources provide aggregated data analysis in readily available charts and maps. These sources will prove to be valuable resources for follow up analyses that may be indicated by the results of the improved Data Sharing System. But we determined that accessing raw data from these sources and pulling that information into our Data Sharing System analyses would be too cumbersome as an ongoing task.

Knowing that data from the U.S. Census, American Community Survey is readily available and easy to match up with our internal fields, we decided to focus on the ACS tables. We reviewed multiple subject tables, including:

<table>
<tr><td>S1701: POVERTY STATUS IN THE PAST 12 MONTHS</td><td>S2101: VETERAN STATUS</td><td>B17001: Poverty Status in the past 12 Months by Sex by Age</td></tr>
<tr><td>B17002: RATIO OF INCOME TO POVERTY LEVEL IN THE PAST 12 MONTHS</td><td rowspan="3">B17001A-F: POVERTY STATUS IN THE PAST 12 MONTHS BY SEX BY AGE A:WHITE ALONE, B: BLACK OR AFRICAN AMERICAN ALONE, C: AMERICAN INDIAN AND ALASKA NATIVE ALONE, D: ASIAN ALONE, E: NATIVE HAWAIIAN AND OTHER PACIFIC ISLANDER ALONE, F: SOME OTHER RACE ALONE, G: TWO OR MORE RACES, H: WHITE ALONE, NOT HISPANIC OR LATINO, I: HISPANIC OR LATINO</td><td rowspan="3">B17020 (A-F): Poverty Status in the Past 12 Months by Age A:WHITE ALONE, B: BLACK OR AFRICAN AMERICAN ALONE, C: AMERICAN INDIAN AND ALASKA NATIVE ALONE, D: ASIAN ALONE, E: NATIVE HAWAIIAN AND OTHER PACIFIC ISLANDER ALONE, F: SOME OTHER RACE ALONE, G: TWO OR MORE RACES, H: WHITE ALONE, NOT HISPANIC OR LATINO, I: HISPANIC OR LATINO</td></tr>
<tr><td>B17022: Ratio of Income to Poverty Level in the Past 12 Months of Families by Family Type by Presence of Related Children Under 18 Years by Age of Related Children</td></tr>
<tr><td>B16009: Poverty Status in the Past 12 Months by Age by Language Spoken At Home for the Population 5+ Yrs</td></tr>
<tr><td>C17002: Ratio of Income to Poverty Level in the Past 12 Months</td><td colspan="2">C18131: Ratio of Income to Poverty Level in the Past 12 Months by Disability Status</td></tr>
</table>

From these tables, we chose fields that most closely matched the demographic characteristics we were including in our data sharing system. We aimed to find data that matched as closely as possible, but often had to accept slight variations. This is to be expected when matching census data with other (case

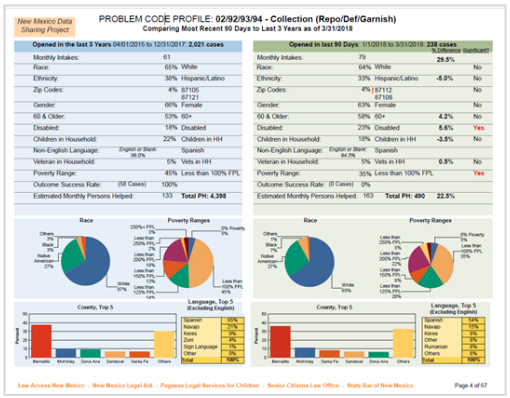
management system) data. We used 5-year estimates because they include data collected from more households and increase statistical reliability by averaging data over 5-year periods. Finally, we downloaded fields from these tables by county and by zip code to match up with our clients' county and zip code data.

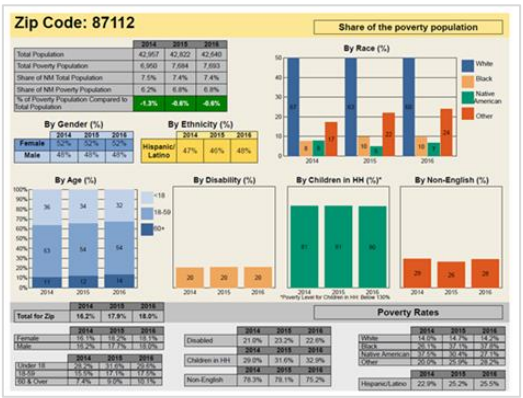
### 5. Linking Internal Data & External Data

As mentioned immediately above, we downloaded external data by county and zip code in order to be able to match up with our analyses of client data by county and zip code. As described in section E. 1. below, restrictions from Pika prevented us from being able to link external data or analysis software directly into the data sharing system. We came up with a solution to this challenge by creating reports with detailed analyses of the external data that include filters allowing for the reports to be limited to zip codes or counties for which a statistically significant change is indicated in the Problem Code Profile Reports.

Because we cannot align the internal data directly to the external data records, we are simply comparing to see where differences appear. We are especially interested in differences that we cannot immediately explain. For example, why are more debt collection clients disabled when the overall data shows that among the eligible population the number of disabled persons is flat? Why is there an increasing share of collections clients from a zip code that is traditionally less poor than other areas? Analyzing the intersection of the internal and external data should help us in at least two ways: 1) to better understand need and unmet need, and 2) to spot trends among eligible people that are already impacting clients or that may impact them in the future.

As an example of how we will be able to link our internal and external reports, the Problem Code Profile for Collections through March 2018 showed that zip code 87112 jumped into the top spot for collections intakes in the last 90 days (whereas zip code 87105 was in the top spot over the last three years). That prompted us to look more closely at zip code 87112. We were able to enter that zip code into our ACS 5-Year Zip Code Data Analysis Report and get more information about 87112, including the fact that though this zip code is less poor than would be expected (the built-in concentration analysis shows that 87112's share of the total population was 7.4% in 2016, but its share of the poverty population was 6.8%), its share of the poverty population has been on the rise (from 6.2% in 2014 to 6.8% in 2016). The report includes several charts and tables indicating the share of the poverty population for various demographic characteristic (e.g., 48% of the poverty population in zip code 87112 is Hispanic/Latino). The bottom section of the report shows poverty rates for various demographics (e.g., In 2016, 75.2% of non-English speakers in zip code 87112 lived below the poverty level).



### 6. Staying Current with External Data

The ACS tables are updated annually and so the external data used to compare with our internal data will need to be updated annually. We knew that it was important to create a process for updating the external data that was as straightforward as possible so that when NMLA staff take over the task from Ms. Perry, there will be clarity about what is required. Thus, we created a process using the U.S. Census API from within Excel that includes details about accessing and formatting the data, which can then simply be refreshed from within our Crystal Report to update the external data analysis.

## E. Project Objective 4: Implement Data Analysis Software

### 1. Tested Open Source Data Analyses Tools

We identified and evaluated multiple potential data analysis tools. Open source options included Microsoft Power BI, RapidMiner, Weka, R, and SaTScan. Non-open sources options that were evaluated include Tableau, Crystal Reports, and Google Fusion Tables. We evaluated the analysis tools for cost, learning curve, data storage and protection, data format requirements, usability on non-numeric data, visuals, and ability to run automatically. Some tools are easy to use and include excellent visuals but have limited statistical analysis capabilities. Other tools include sophisticated analyses capabilities but have limited visuals and may be difficult for users to learn. Some are free, and some have costs. We aimed to select the tool that balances analysis prowess with ease of use, accessibility of results, and affordability.

We discussed the potential for attaching one of these tools to the NM Data Sharing System with Aaron Worley from Pika and he was concerned about ongoing support for the tool, as it will not be possible for him to manage it. With this information, and based in large part on cost and learning curve issues, we decided that the best solution is to use Crystal Reports since NMLA is already using it to analyze Pika data. The Problem Code Profile Reports using Two Population Difference of Proportion tests described above were created using Crystal Reports. Because of staff turnover at NMLA, Ms. Perry has agreed to download the NM Data Sharing System data quarterly, run the Crystal Reports on those data, and send the reports to the partner organizations through 2018. She has further agreed to train NMLA's newly hired Grants Manager/Data Analyst so that she can take over this task in 2019.

### 2. Learned to Use Tool with System Data

Ms. Perry is an expert user of Crystal Reports and has been using the software to analyze legal aid data for more than eight years. In fact, she has created a Crystal Reports training guide specific to NMLA that uses NMLA Pika examples exclusively and has been training our new Grants Manager/Data Analyst on Crystal Reports for the last six months. That general Crystal Reports training will continue throughout 2018 along with training specific to the reports created under this project.

The Crystal Reports we created for this project include a series of complicated statistical formulas for Frequency Comparisons and population estimates, standard errors, and z-scores to perform the Two Population Difference of Proportion Test. Now that those formulas are created and tested, the reports can simply be re-run with fresh data whenever we, or any of our partners, would like access to the analyses. The most complicated parts of the process will be accessing and preparing the data for analysis rather than running the Crystal Reports themselves. Ms. Perry created detailed instruction guides for preparing internal data downloaded from the data sharing system, linking that internal data to the Crystal Reports, using the U.S. Census API to access external data, and linking that external data to the

Crystal Reports. The combination of those instructions and Ms. Perry's training will ensure that we are able to continue running these reports quarterly and distributing them to our partners as planned.

### 3. Began to Use Data Tool to Foster Collaboration

The project partners also began to use the data tool to identify substantive legal areas where multiple agencies could use the data to foster improved responses, issue-oriented organization-specific strategies, and issue-oriented collaboration and coordinated strategic advocacy among the partner organizations. We admittedly did not get as far down that road as we had hoped. As indicated in the report, the challenges with resolving data variations across the partner agencies, completing the definitional mapping between data fields used by the different agencies, and working through the steps needed to decide which statistical comparison tests were most workable and effective for the project, all took significantly greater time than we had anticipated. Staff turnover at NMLA's grants manager position, which handles most of our data analysis tasks, also slowed us down.

We have a new grants manager on board, however, who will use the data tool to generate monthly reports to analyze new trends and issues. We will also use the reports internally, to be reviewed by our Litigation Director in collaboration with the leaders of NMLA's four statewide practice groups – family law, housing law, consumer law, and economic security (includes government benefits programs). Our Litigation Director will coordinate with the partner agencies to develop strategies for responding to trends and issues that affect clients of more than one agency.

Some examples of initial issues that we are looking at include consistently high correlation between clients who are disabled and consumer debt cases; and an increasing number of clients over age 60 who are seeking help with cases involving custody or guardianship of grandchildren. The new data tool will let us also more accurately factor in variations attributable to local demographic and geographic data. This capacity in turn will let us determine whether a targeted local or regional approach will be more effective than a statewide approach for legal problems that may be impacting one community or client group more than others.

## F. Project Objective 5: Use the Tool to Better Serve Clients & Foster Collaboration

### 1. Developed Analyses of Data Relationships

The analysis tool went through 52 versions involving significant testing before being finalized. Before creating what has become our final version, Ms. Perry conducted several rounds of analyses designed to uncover various trends in the data and several rounds of additional analyses designed to test the validity of the data definitional mapping results. For much of the report testing period, the data uploading and mapping issues were not fully resolved.

Once the data mapping issues were mostly resolved, Ms. Perry created an early version of the analysis report that showed subject area profiles and analyzed the original fields collected in the data sharing system. Realizing that the substantive area level (example: "family law") was not detailed enough to allow the partners to uncover important data changes and relationships, she created profiles at the specific legal problem level (example: "divorce"). Multiple updates to the analysis report occurred over the course of several months due to extensive testing, including: preventing overlap of cases between the 3-year period and the 90-day period, excluding blank data, the addition of triggers that limit the results to those with statistically significant changes, the creation of a detailed formula to map problem

codes from two different problem code fields, the creation of formulas that addressed fields with differing formats (number and text), the addition of new data fields and the associated statistical tests, and several formatting changes. Often the report testing uncovered additional edits required for the field mapping within the data sharing system, which had to be resolved, data re-downloaded, and analyses re-tested.

The resulting Problem Code Profile Reports consist of two columns on the first page representing data from the last three years (blue column on the left) and data from the last 90 days (green column on the right). Within the boxes at the top are listed various case and client fields and information about how their proportions are different in the two time periods and whether that difference is statistically significant (using the Two Population Difference of Proportion Test).

The Problem Code Profile Reports also include the following additional analyses:

1. Pie and bar charts and a table are used to provide additional information about Race, Poverty Ranges, County, and Language.
2. Monthly Intakes: While the total intakes for the 3-year period and the 90-day period appear at the top of the report, we added the average monthly intake numbers for those same periods with an indication of the percentage difference between the average monthly intake numbers for each time period.
3. Outcome Success Rate: Our field mapping procedures already combined our organizations' varying outcome measures into broad categories, which group positive outcomes into a category called, "Won/Hearing Won/Settled Favorably" and negative outcomes into a category called, "Lost/Hearing Lost/Settled Unfavorably." In the future, as organizations add more substantive outcome measures, we could expand the categories to be more detailed. The Outcome Success Rate is calculated by dividing the number of wins by the number of wins and losses, for each legal problem code. Additionally, we included a count of the cases with outcomes so that the success rate can be reviewed based in part on the number or cases involved (100% success rate with only one case is not informative, whereas 95% success rate with 150 cases is useful information).
4. Estimated Monthly Persons Helped: Because one organization does not provide persons helped data, we calculated an average number of persons helped per case by legal problem from among the cases with persons helped data and then multiplied that average by the total number of cases with each legal problem to get an Estimated Total Persons Helped. We simply divided that by the number of months to get the monthly estimate and provided an indication of the percentage difference between the average monthly numbers of persons helped for each time period.
5. Each legal problem with at least 50 cases in the most recent 90 days includes a page with opposing party, court, and judge information:
    a. Top 5 most frequently occurring opposing parties: Note that for some legal problems most, if not all, of the opposing party data is redacted.
    b. Outcomes by Court: The count of cases with outcome wins and outcome losses are listed by Court. Note that the Court field is one that required a lengthy formula to sort the various spellings for different courts.
    c. Outcome by Judge: The count of cases with outcome wins and outcome losses are listed by Judge.

6. Finally, for every case or client field that shows a statistically significant change, cross tabs appear that provide additional information about the relationships between the field with the statistically significant change and other case and client fields. For example, if a particular legal problem code shows a statistically significant change in the Disabled field, cross tabs will be generated comparing the Disabled field to other fields such as race, ethnicity, age, gender, children in the household, language, veteran in the household, and poverty range.

Disabled Data Relationships

02/92/93/94 - Collection (Repo/Def/Garnish) (2,021 cases)
Opened in the last 3 Years: 04/01/2015 to 12/31/2017

| | Disabled | Others | Total |
|---|---|---|---|
| White | 53% | 38% | 41% |
| Not Entered | 37% | 38% | 38% |
| Others | 9% | 24% | 22% |

| | Disabled | Others | Total |
|---|---|---|---|
| Hispanic | 37% | 38% | 38% |
| Others | 63% | 62% | 62% |

| | Disabled | Others | Total |
|---|---|---|---|
| 60 & Older | 77% | 48% | 53.3 |
| Others | 23% | 52% | 46.69 |

| | Disabled | Others | Total |
|---|---|---|---|
| Female | 61% | 67% | 66% |
| Male | 39% | 33% | 34% |

| | Disabled | Others | Total |
|---|---|---|---|
| Children in HH | 5% | 26% | 22% |
| No Children | 95% | 74% | 78% |

02/92/93/94 - Collection (Repo/Def/Garnish) (238 cases)
Opened in last 90 Days: 1/1/2018 to 3/31/2018

| | Disabled | Others | Total |
|---|---|---|---|
| White | 49% | 39% | 41% |
| Not Entered | 36% | 35% | 35% |
| Others | 15% | 26% | 24% |

| | Disabled | Others | Total |
|---|---|---|---|
| Hispanic | 35% | 32% | 33% |
| Others | 65% | 68% | 67% |

| | Disabled | Others | Total |
|---|---|---|---|
| 60 & Older | 73% | 53% | 58% |
| Others | 27% | 47% | 42% |

| | Disabled | Others | Total |
|---|---|---|---|
| Female | 60% | 64% | 63% |
| Male | 40% | 36% | 37% |

| | Disabled | Others | Total |
|---|---|---|---|
| Children in HH | 5% | 22% | 18% |
| No Children | 95% | 78% | 82% |

New Mexico Data Sharing Project

Statistically Significant Changes Between the the Last 3 Years (04/01/2015 to 12/31/2017) and the Most Recent 90 Days (1/1/2018 to 3/31/2018)

| | Race | Ethnicity | Zip | 60 & Over | Age Range | Gender | Disabled | Children | Vets | Poverty |
|---|---|---|---|---|---|---|---|---|---|---|
| 01/85 - Bankruptcy/Debtor Relief | | Yes | | | | | | | | |
| 02/92/93/94 - Collection (Repo/Def/Garnish) | | | | | | | Yes | | | Yes |
| 03/96 - Contracts/Warranties | | Yes | | | | | | | | |
| 09/99 - Other Consumer/Finance | Yes | | | | | | Yes | | | |
| 29 - Other Employment | | | | | | | | | | Yes |
| 31 - Custody/Visitation | Yes | | | | | | | | | |
| 32 - Divorce/Separ./Annul. | | | | | | | | | | |
| 37 - Domestic Abuse | | | | | | | | | | |
| 38 - Support | Yes | | | | | | | | Yes | |
| 38/49 - Other Family | | | | | | | | | | |
| 42 - Neglected/Abused/Depend. | | | | Yes | | | Yes | Yes | | |
| 44 - Minor Guardianship/Conservatorship | | | | Yes | Yes | | Yes | | | |

In addition to the Problem Code Profile Reports, we also have the ACS 5-Year County Data Analysis and ACS 5-Year Zip Code Data Analysis reports described in section 4.D.5. above and the Summary of Statistically Significant Changes report that shows all problem codes with statistically significant changes in various demographic characteristics so that we can look to see which problem codes are showing similar changes and might, therefore, be related in some way.

## 2. Trained Partners

On October 23, 2017, we introduced our new Problem Code Profile Reports to our partners and trained the partner representatives on how to read and understand the report results. We explained the Two Population Difference of Proportion Test, along with the charts, graphs, opposing party information, outcomes by court, and outcomes by judge cross tabs.

The data used for the training covered the period between October 1, 2014 and September 30, 2017, data for the 3-year period (October 1, 2014 to July 2, 2017) and the 90-day period (July 3, 2017 to September 30, 2017). There were some problem codes with statistically significant changes in proportions in the 90-day period compared to the 3-year period and some additional problem codes with no statistically significant changes, but with 50 or more cases opened. The partners determined that they wanted to receive the report results for both those with statistically significant changes and those with at least 50 cases.

They also requested a version of the report that separates out cases from Bernalillo County (Albuquerque metro area). Given that one-half of New Mexico's population lives in Bernalillo County, this makes it likely by definition that Bernalillo will always produce the numerically highest results for any county searched. By separating Bernalillo County from the aggregate results, therefore, the rankings of other counties within any data selected for a query will become more meaningful.

Project partners also requested customized report versions that are unique to each organization. Following the training session, we received the following feedback from Kathy Heyman at SCLO: "You did a great job not only with compiling all the data but also with your presentation. Thank you so much.

Also, the data you accumulated and reported on that pertained to SCLO clients was spot on. We have seen a big increase in landlord tenant cases with mostly female clients."

We sent updated versions of the reports to our partners for data through December 31, 2017 and reviewed those reports with them in February. In addition to some small formatting changes, we decided that adjustments needed to be made to our client age and age over 60 fields to ensure consistency. Those changes were made, and final versions of the reports were distributed for data through March 31, 2018.

### 3. Surveyed Partner Organizations

Because our data mapping challenges delayed our ability to finalize the analyses, partners have thus far just received three rounds of the reports (one for data through September 30, 2017, one for data through December 31, 2017 and one for data through March 31, 2018). We are optimistic about how we and our partners will use the analyses individually and collaboratively. A survey distributed to the partners shortly after they received the March 31, 2018 reports revealed that all partners plan to increase their analysis of case management data because of our reports. Four out of five partners will use the Problem Code Profile Reports to better understand clients and the fifth partner may use them.

Three out of five partners found the profile reports easy to understand, while two partners found some parts easy to understand and some parts difficult to understand. All partners found the zip code and county analyses reports either easy to understand or very easy to understand and two partners report that they will use those external reports to better understand needs in the community, while three partners report that they may use them.

We asked the partners to rank parts of the Problem Code Profile Reports in order of helpfulness. The demographics tables at the top of the report indicating proportions and statistically significant changes were ranked most helpful, followed by the demographics charts and graphs at the bottom of the first page. Next, the opposing party tables and tables showing relationships between demographics were tied, and the outcomes by court and judge tables were selected last. We were not surprised at the order, largely because the items at the bottom of the ranking are the ones for which very little data are collected or uploaded to the system. As our partners begin gathering more outcome data and reporting more opposing party, judge, and court data, we believe these sections will become much more informative.

Finally, we are pleased that all partners report being satisfied or very satisfied with our process and satisfied or very satisfied with the reports developed from the process.

## G. Project Objective 6: Deliverables: Sharing our Findings

### 1. Developed Guidelines

We learned important lessons about collaborating with partner organizations and especially the complicated process for definitionally mapping fields across organizations. We also identified creative solutions when faced with unexpected software challenges. These things, along with tips about internal data and external data sets, analyses methodologies, partnering with data analysis experts, and reporting suggestions are all included in the guidelines we shared at the January 2018 ITC Conference that we hope will be used by other legal aid programs in the future to collaborate with partner agencies and aggregate data to better understand clients and their legal needs.

### 2. Publicly Presented Results

Ed Marks and Rachel Perry presented our and findings and guidelines at the January 2018 Innovations in Technology Conference in a session titled "Powerful Public Data Visualizations at Your Fingertips." While the session had a particular focus on accessing external data, we presented many details about the project, including our multi-organizational partnership, data mapping tasks, the Two Population Difference of Proportion test, our Problem Code Profile Reports and the related external data zip code and county profiles. We provided very specific tasks for setting up and using the U.S. Census API for accessing American Community Survey data, including directions on where to find legal aid-relevant ACS data tables, a list of recommended ACS fields for comparison to legal aid data, and step-by-step instructions for using the API in Excel. We also shared other external data resources that we discovered during our project including DataUSA and the Kids Count Data Center.

## IV.a. Information for Multiyear or Multiple Projects

This project builds off the TIG project in which we originally created the New Mexico Data Sharing System (#12018).

### 1. Improvements to Existing Data Sharing System

At the end of that original project, we identified future improvements that we hoped to accomplish, including adding new fields to the system, adding new filters to the system, setting up automatic email alerts, and investigating how to help partner organizations compare aggregated system data to their own internal confidential data. Based on extensive review, we identified and implemented those and additional significant improvements, including updating the Trends tab, creating a new Problem Codes Trends tab, and incorporating a method for downloading the entire database. A few of the improvements changed during this phase, such as the partners deciding not to upload confidential data and rather to run comparisons of system data to their own internal data outside of the system. Also, after receiving the email alerts for about ten months, the partners decided that they were too frequent and not informative enough and that they would rather receive the new analyses reports on a quarterly basis.

### 2. Data Compatibility Issues

We started to address the data compatibility issues in the original project but were not able to fully resolve them until the current project. While we knew there would be some field mapping issues, we did not foresee the extensive nature of the field mismatch across organizations, nor the significant amount of time it would require to write complex coding and formulas to ensure consistent categorization of and apples-to-apples comparisons of intra-organization data. Some organizations define fields differently and/or have different answer options, requiring complex programming to map fields by definition so that they can be aggregated for analysis. In particular, race, ethnicity, legal problems, outcomes, language, and close codes required complex definitional mapping. Additionally, some fields that allow users to enter textual answers are riddled with spelling issues or issues around how data are identified. For example, the County field has many misspelled entries, making it difficult to compare county-specific data. Another example is the Court field, for which some entries are spelled out (Fifth Judicial District Court), some use numbers (5th Judicial District Court), and some use abbreviations (5th Jud. Dist. Ct.). Formulas were written to address these issues.

### 3. Analysis of Multiple Issues & Multiple Factors Simultaneously

Our original version of the data sharing system showed increasing or decreasing intake trends for legal problem codes, but without some significant degree of manual filtering of data and individual interpretation of the resulting data sets, it did not automatically identify simultaneous changes among multiple legal problems nor did it include any analysis of demographic characteristics. The reports created through the current project do just that and provide information about changes in the eligible client community that might be impacting the intake trends.

### 4. Practical Uses for Analyses

Having greatly improved the quality of the data and the scope of the analysis generated by our data sharing system, we are on the cusp of being able to make more strategic, proactive, and informed decisions about advocacy efforts, case acceptance, and outreach.

## V. Factors Affecting Project Accomplishments

We are pleased with the results of this project. Because of our efforts, we have developed tangible tools: the internal data Problem Code Profile Reports and the Summary of Statistically Significant Changes and the external ACS 5-Year Zip Code Data Analysis Report and ACS 5-Year County Data Analysis Report.

### 1. Success Because of Partnerships

The most significant factors influencing our success were our partnerships with data analysis experts and the partnership among legal service organizations. The data analysis experts helped us conquer mountains of internal and external data, address the variations in field definitions across organization by definitionally mapping fields, identify the most relevant data questions, formulate fitting analyses, identify the best cost-effective software, and communicate our findings in user-friendly formats. Our legal service partners not only agreed to share their data but also provided valuable input as we formulated our analyses and developed user-friendly and informative reports. Still, we did run into a few challenges, which are listed below and the solutions to which are described in the next section.

### 2. Data Challenges

As discussed previously, while we knew there would be some field mapping issues, we did not foresee the extensive nature of the field mismatch across organizations, making comparisons of intra-organization data difficult and sometimes impossible at the start of the project. Some organizations define fields differently and/or have different answer options. These were particular problems for the race, ethnicity, legal problems, outcomes, language, and close code fields. Additionally, some fields that allow users to enter textual answers are riddled with spelling issues or issues around how data are identified. For example, the County field has many misspelled entries, making it difficult to compare county-specific data.

### 3. Staff Turnover Challenges

Another challenge arose out of turnover in the Grants Administrator/Data Analyst position at NMLA. Because of the turnover, Ms. Perry was not able to fully train the person in that position regarding the structure of the reports and the steps required to update the data and reports every quarter.

### 4. Software Challenges

One of the organizations (Law Access) experienced a significant coding issue with the uploader software that provides data from their version of Pika to the data sharing site in the middle of the project. Also, once the project was well underway, we learned that Pika could not oversee the install or maintenance of knowledge management tools within the data sharing website as we originally hoped could be done.

## VI. Strategies to Address Major Challenges

Fortunately, we were able to develop solutions to every challenge we faced.

### 1. Data Challenges

Data challenges are to be expected with these types of projects. Still, the extent of the data challenges was unexpected. Fortunately, Ms. Perry was able to write complex coding and formulas to address the field mapping issues, thereby ensuring consistent categorization of and apples-to-apples comparisons of intra-organization data. Additionally, she wrote code to categorize answers to fields that allowed for unlimited write-in text answers. She was careful to test and re-test her coding and formulas multiple times before incorporating them in the analyses reports.

To address the two legal problem fields used by SCLO, Ms. Perry wrote and tested programming to map problem codes by their meaning within the data sharing system, and after months of trying to get this to work within the data sharing system, determined that we had to upload both sets of problem codes and implement the coding to map the fields in Crystal Reports, rather than within the data sharing system. An example of the legal problem coding in place within Crystal Reports to aggregate cases with Medicaid issues (if LSC Problem Code = 51-Medicaid OR SCLO City Problem Code is one of [25-Medicaid, 28-Medicaid Waivers, 29-Institutional Medicaid Eligibility], then use problem code label “Medicaid”).

This field mapping process required many rounds of testing to ensure that it is working. In fact, we provided the first field mapping definitions to Pika in June of 2016 and had to download and test the data, run results by our partner organizations, work with Pika on edits and corrections, and develop mapping formulas in Crystal Reports through September of 2017.

### 2. Staff Turnover Challenges

NMLA’s new Grants Manager/Data Analyst started in December of 2017 and has been receiving Crystal Reports training and support from Ms. Perry since she started. That general Crystal Reports training will continue throughout 2018 along with training specific to the reports created under this project. In the meantime, Ms. Perry has agreed to update the report through 2018 to ensure that NMLA’s Grants Manager/Data Analyst has sufficient time to become comfortable with Crystal Reports in general and the reports for this project specifically.

### 3. Software Challenges

As a temporary fix to Law Access’s problems with the uploader software, the organization provided their data in raw form throughout most of 2017, which Ms. Perry combined with the other organizations’ data from the data sharing site in Microsoft Access, which was then brought into Crystal Reports for analysis. Extensive work was required to get this organization’s data categorized and mapped to the other organizations’ data each time data were downloaded. She wrote extensive instructions on the steps required to aggregate these data files so this temporary solution could be in place for as long as it

took to resolve the uploader issue. Fortunately, the uploader coding issue was resolved in November of 2017 so we could revert back to gathering all data from the data sharing site.

Finally, because of Pika's inability to host analysis software within the data sharing site, we decided that the best solution was to have NMLA use Crystal Reports, which it has been using for years already, to analyze the aggregate data from the data sharing site.

## VII. Major Lessons and Recommendations

There is enormous potential for legal aid organizations to significantly improve their client service effectiveness and to proactively address emerging by collaborating with organizations providing similar services to the similar people. Yet, there are some potential challenges for which any effort at collaboration needs to prepare:

- The most significant challenge is related to definitionally mapping fields to allow for apples-to-apples comparisons across organization. It is essential that there is a clear understanding among all the partners as to exactly what each shared field means. This task can be very time consuming, so it is important to allow for sufficient time to work through it.
- It is important to engage with partners regularly and to seek their input and encourage their buy-in.

Similarly, there is enormous potential for legal aid organizations to gain a better understanding of their clients and the complicating factors that impact their legal needs by employing more sophisticated statistical analyses than has been the norm. Yet, some planning steps are important before applying any analysis method:

- Getting assistance from data analysis experts right from the start is essential so that they can steer you towards the best, most applicable analysis methods that will work with your specific data. We started out wanting to apply multi-variable correlation analyses but learned that our data issues prevented us from being able to use that specific kind of analysis. Our experts were able to identify an alternative method that still provided very useful information.
- It is important to have a staff person with data analysis skills who can keep the analyses going. We want the analyses created during this project to continue and we want to be able to figure out new and impactful ways of using the analyses. That requires an in-house staff person to carry on the work of the data analysis experts.

Finally, based on the results of our analyses, we have reached some broad conclusions:

- When reviewing data analyses, find the results that are surprising or that cannot be easily explained. That is usually where there is something going on for clients that needs attention.
- Comparing internal case management data with external data about the eligible population is an important way to understand changes in potential clients' needs.