To: Terrence Woods and Kathryn Helm
From: Office of Reporting, Research, Analytics, and Implementation, ODHS
Subject: Oregon's Data Strategy
Date: December 15, 2020

The Office of Reporting, Research, Analytics, and Implementation (ORRAI), positioned under the Director's Office in ODHS, serves to meet research and analytical needs both within and across the five ODHS programs, as well in partnership with other agencies such as DOC, OHA, ODE, and others.

Crafting a strategy that is general enough to apply across the breadth of Oregon government, but specific enough to provide meaningful guidance to the data owners and consumers of the disparate systems, is most definitely a challenge. We applaud the thought and careful work of the current draft, appreciate the opportunity to comment on the emerging Oregon Data Strategy, and look forward to a productive future with all our state partners as we implement the strategy.

ORRAI's primary concern is that the proposed strategy plans well for how Oregon should *treat* its data, while providing less clarity, at least for now, for how Oregon should *leverage* its data.

How Oregon should treat its data

The first of the strategy's themes—governance and effective management—is well developed. The current text of the strategy provides direction for surveying and prioritizing the State's data resources, for centralizing their storage and maintenance, formalizing documentation of the contents, context, and meaning of data elements, and building both the infrastructure for and the cultural underpinnings of transparency, openness, and sharing.

We are pleased to see that GIS initiatives, Oregon's Framework Program and GEOHub, have been included as the model for expanding data assets to the enterprise and for secure sharing with community and stakeholders. Spatial data, and the specialized analysis provided by Oregon's geospatial information systems professionals, should be lauded, supported, and enhanced in the data strategy. Their value in Oregon's dual 2020 crises, the COVID-19 pandemic and damaging wildfires, reveals how valuable spatial data and spatial data experts are to an effective and nimble government.

The strategy should encourage increasing the richness of key demographic variables

ORRAI agrees with the strategy that demographic data should be enhanced, as the implementation of REAL-D will support. That said, we would like to see the state stress that agencies begin to document gender identity with variables that go beyond a binary conception of sex.

The strategy should call for an effort to recognize family relationships in Oregon data

Additionally, ORRAI would like to see the state support efforts to identify individuals as part of families—with all due cultural sensitivity, understanding, and humility about what constitutes a family. Much of state research could be improved by appropriately acknowledging that individuals are not independent of everyone else, and that Oregonians live their lives as members of families. Without carefully constructed family identifiers in our data though, these analytic efforts are inhibited.

The strategy should include the creation of a state-level master client index

A state-level master client index is critical foundational infrastructure, necessary to many of the strategy's goals: centralizing and documenting state data, encouraging data sharing, dissemination, and openness, and better securing the privacy and anonymity of the Oregonians on whom our data focus. Creation and maintenance of a master client index is not trivial, but failing to create a unique, person-level identifier across

the enterprise's data systems risks balkanization and siloing—some of the very same poor outcomes the proposed strategy seeks to remedy. In our present state, without a central, shared client index, agencies and programs can only share information and connect records using personally identifiable details—names, dates of birth, addresses, social security numbers, etc. In addition to improved privacy protections for Oregonians, the creation of a master client index would also greatly enhance the state's ability to leverage disparate data systems in service of common goals.

How Oregon should leverage its data

The strategy supports regular reporting

The strategy can improve Oregon's abilities to provide regular, routine, standardized reports—the necessary documentation to interrogate the state's stewardship of resources, evaluate whether it is meeting its commitments to positive outcomes for Oregonians, and advance its partnerships with businesses, non-profit service providers, cultural institutions, and everyone working to make Oregon vibrant, thriving, healthy, and safe. Oregon's regular reporting could benefit, and much could be automated, from the strategy's inventorying and linking together the state's most valuable data resources, documenting the context and provenance of the raw data elements, and centralizing cleaning, standardizing, deduplicating, and other key processing of data sets.

The strategy should increase the state's research and analytic capacity

Government data should not be limited to just reporting. Research and data analytics can help government move beyond just client counts and service metrics. Used appropriately, they can inform both the array of individual-level decisions made in the field as well as upstream decisions from leadership to change the processes and trajectories of entire agencies.

Research can reveal disparities in decisions, services, and outcomes by region, race, ethnicity, gender, disability. It can help reveal putative risk or protective factors for important personal or communal outcomes. It can help distinguish more effective services and initiatives from less effective.

Data-informed analytic tools can leverage data to anticipate future outcomes. They can match clients' needs to state services. They can support staff and leadership throughout the enterprise, faced with fraught, consequential decisions.

While centralizing data storage, documentation, and processing can make for more efficient reporting, many agencies lack the research and analytic expertise to leverage data productively, no matter how clean or available the data may be. The data strategy should call for a centralized resource for cross-agency research and advanced analytics, including but not limited to machine learning or predictive algorithms, and development of algorithmic fairness techniques. Staff in this centralized resource would partner with agency research analysts, drawing on the latter's keen knowledge of the provenance and meaning of the data elements within their agencies and as experts on identifying productive use cases. These partnerships would develop the cross-agency research, and decision-support tools, for creating more data-informed decisions throughout the enterprise.

This call for increased research and analytic capacity should not neglect geospatial information systems. At ODHS/OHA, we use GIS and geospatial data in a variety of areas. They help improve data visualization, through maps, dashboards and surveys. Spatial analyses reveal trends and relationships through efficient geoprocessing. Geospatial data can be incorporated into statistical modeling, as ODHS has been exploring using spatial features to aid predicting areas of the state with higher instance of child abuse. Additionally, spatial data and GIS are often key to bridging disparate data sources. Many program data sources are not well linked through common identifiers and can only be joined for exploration, study, and understanding through their spatial relationships. Spatial data and GIS can expand our analytic capabilities in multiple dimensions.

The strategy should balance the need for centralization with the need for individualization

The strategy can help create a more standardized data infrastructure and should advocate that the state invest in enhanced research and analytic capacity in all of the contributing agencies. At the same time, though, the strategy should allow for relative autonomy in the research and analytic efforts of the agencies.

As we have written above, we advocate for a centralized research and analytics resource for the state. Our hope is that this centralized resource would be a partner with autonomous agencies, helping them accomplish their research and analytics needs, rather than driving those efforts top-down. The agencies have the business expertise to know where centrally designed tools might best be applied, and which data elements can best benefit efficiencies and transformation within the agency.

Relatedly, the strategy should make clear that the centralization and standardization of the state's data should be a net benefit to the speed, efficiency, and nimbleness of both agency and enterprise-level research and analytics efforts. Governance at the level of the state data strategy should be geared toward making it easier for agencies to use data to improve their business, rather than slowing innovation and improvement by requiring new layers of required permissions and processes for agencies to use their own data.

The strategy should provide a balanced view of predictive analytics

The current strategy mentions research and analytics mostly in passing. When it does provide further detail, it is almost uniformly negative: focusing on algorithms' potential for inequitable benefits and harms and listing some high-profile failures in conducting research in Oregon or implementing predictive tools in other jurisdictions.

ORRAI advocates for research and decision-support tools, including predictive algorithms. However, we assert that they should be carefully designed and thoughtfully implemented as part of data-informed decision processes, with an equity lens approach. These tools should be optimized and evaluated not solely for traditional metrics like accuracy—which may well automate the biased decisions of the past—but also for fairness, across a variety of definitions of fairness.

Oregon has reason to be proud when it comes creating data-informed tools, while acknowledging that the data on which we build those tools carries the biases of the past: troubling differences based on race, gender, ethnicity, tribal affiliation, class, etc. resulting in different data histories due to over- or under-monitoring, different paths through government involvement driven by unfair decision processes, and differing outcomes and experiences due to systemic inequities in society. In particular, ORRAI has done ground-breaking work addressing these biases, and is leading the nation in incorporating algorithmic fairness in tools based on government data, reducing the possibility that the mistakes of the past are perpetuated into the future. You can find a preprint publication by ORRAI researchers Jordan Purdy and Brian Glass on fairness correction in a Child Welfare algorithm from https://arxiv.org/abs/2010.12089.

The state data strategy should rebalance its presentation of tools like predictive analytics and include the good work that ORRAI and others are doing to advance data justice and improve fairness, when using data to improve government work.

The strategy should highlight the importance of implementation professionals to a data-informed culture

The state data strategy focuses on important infrastructure on the "back end" of the data, where data is prepared and made accessible to data reporting, research, and analytic staff. This is the physical and organizational structure around what data is taken up, how it is stored, standardized, and processed.

There is little coverage in the strategy, though, of important infrastructure on the "front end," where the products from work by reporting, research, and analytic staff are translated into useful information for field

staff, management, and leadership. This is the organizational structure around how data products are turned into changes in processes, into better decisions, and into improved outcomes. This necessary work is the domain of implementation professionals.

Solid implementation of data products contributes to many of the goals of the state data strategy. The strategy rightly worries about automated decision tools. In ODHS work, though, our implementation staff take great care to ensure that consumers of data understand that ours is a data-informed, not data-driven culture. Tools created from historical data can be a great boon to individuals faced with difficult and important decisions, those deciders are trained that the data-based tools are just one source of information to bring to bear on these decisions, and that the information they provide should be combined with other sources of information, as well as professional discretion, in forming the final decision.

Additionally, the strategy has the laudable goal of improving the data-literacy of the government workforce. However, as beneficial as general training and resource libraries about being a general consumer of data might be, it is likely more valuable to gain proficiency and trust in the specific data products that staff will actually be using on the job—on the reports, dashboards, predictive tools, etc. that come from the state reporting, research, and analytics shops. Learning and acceptance can be facilitated when topics move from the abstract to the concrete, and we might expect that working with implementation professionals to understand a tool the improves on the work of staff in program, can lead to better gains in data-literacy than generalized coursework.

In these ways, implementation professionals help build bridges between researchers and the staff consuming the researchers' work. But, implementation professionals, with their expertise in communicating about complex, data-related topics to a variety of audiences, can also help build bridges between research and community. Implementation professionals can facilitate bringing the community voice to researchers—as ORRAI has demonstrated with, for example, The Klamath Tribes in helping foster a meaningful research agenda around The Klamath Tribes' interactions with Child Welfare. Implementation professionals can also facilitate communicating the results of research and analytics work to community members and other stakeholders, and provide the openness and transparency called for in the strategy, and necessary to build trust in how the state is using data.

The state data strategy should call for the creation, expansion, and support of implementation teams in each agency, to support the thoughtful adoption and use of data-based tools by staff, and to help the two-way communication between research and our Oregon communities.

Thank you for this opportunity to provide feedback on the Oregon Data Strategy. ORRAI looks forward to a productive partnership with the Chief Data Officer and we are happy to offer our assistance at any time.