

Chief Research Informatics Officer (CRO) vision statement

Melissa Haendel, PhD

January 21st, 2021

Executive Summary

The University of Colorado Anschutz Medical Campus (CU-AMC) comprises a large and diverse biomedical research community. Conversations with approximately thirty senior leaders, faculty, and staff from across CU-AMC, Children's Hospital, and UCHealth have revealed a fragmented suite of institutional data, analytical, and personnel assets in informatics and infrastructure. These assets hold tremendous potential for CU-AMC, regional hospitals, and partnering organizations to realize the promise of precision medicine, learning healthcare, and global research prominence.

The CU-AMC Chief Research Informatics Officer (CRO) and associated office will empower CU-AMC with the leadership it needs to thrive in an increasingly data-driven world and will coordinate the use of information and information systems to transform biomedical research and improve quality and efficiency. As a member of senior leadership, the CRO will provide strategic direction and expertise necessary to maximize the impact of CU-AMC's research and data assets, recognizing the need for careful socio-technical integration and sound governance. The CRO office will focus on integrating analytic and data capabilities to maximize the impact of CU-AMC's data assets and aid in the realization and advancement of the following key business differentiators: 1) Learning healthcare; 2) health system and research operational success; and 3) workflows and infrastructure to promote new translational discoveries and their implementation. The office will foster a data ecosystem that maximizes the return on investments in personnel, data production, and informatics infrastructure taking into account the needs of a diverse array of organizational partners, patients, and communities.

Biomedical research and clinical care require increasingly sophisticated and well coordinated collaborative analytics that only dedicated effort can provide. Progress in disease prevention, early detection, diagnosis, prognosis, and treatment is increasingly informed by analysis and interpretation of multi-modal data, modern analytical techniques such as machine learning and artificial intelligence, and by integration of public and local data and knowledge sources. Hypothesis evaluation has shifted from challenges in data generation to addressing computational challenges to process and analyze the data; using the many relevant data and knowledge sources requires computational skills, computing resources, quality data management, and a diverse and collaborative workforce - not unlike a care team. Institutions that have invested in their informatics infrastructure, expertise, training, and data governance are experiencing greater efficiencies, innovations, financial outlooks, and importantly, improved care and patient outcomes. While a few programs have made advances in coordinating informatics and data assets at CU-AMC, many independent faculty and staff with data science skills across campus are potentially isolated from complementary and synergistic expertise, computational resources, and data. Data access, sharing, and analytics at CU-AMC is challenging, resulting in duplication of effort, extra costs, and inefficiencies. There exists a significant need for leadership to understand the current capacity and gaps in the informatics landscape and to foster strategic planning for a wide variety of institutional priorities. Put simply, the potential of broadly integrated data and information in combination with coordinated healthcare data management, academics, research services, and teaching to effectively address biomedical research problems requires campus-wide strategic leadership.

CRIO Mission

The CRIO will serve as a primary liaison between biomedical researchers and information technology (IT) teams, analytics initiatives, and data science education efforts on campus. In doing so, the CRIO office will assume a critical role as central brokerage, identifying gaps between existing resources, suggesting further investments, driving innovation, and addressing inequities. The CRIO will act as a strong, well-informed advocate for researchers and stakeholders, seeking especially to understand and encourage potentially reluctant members who are key to the mission. To achieve success in its mission, the CRIO office will work in federation with existing efforts - implementing a distributed ("snowflake") model of organization that will both solidify relationships with the wider community while empowering each participant to maintain autonomy in how they direct their work. The following mission areas have been identified in discussions thus far:

- **Collaboration.** Serve as the nexus and force-multiplier to facilitate team science among investigators with computational and analytical expertise. Facilitate collaboration with other informatics and data science communities and groups throughout the state, country, and world.
- **Governance.** Coordinate and evolve governance processes for data resources and their management across CU-AMC and in collaboration with key partners such as Children's and UC Health.
- **Informatics assets.** Perform ongoing landscape inventory and evaluation of existing data assets, computational resources, expertise, and programs.
- **Innovation.** Liaise with investigators to support informatics innovation, translation, and dissemination in coordination with CU Innovations, the CCTSI and other campus partners. Aid the creation of public-private partnerships and business innovation.
- **Regulatory and security.** Provide recommendations to balance risk, mitigation, and research and care goals in coordination with regulatory and security partners and considering patients as key stakeholders.
- **Learning healthcare.** Assist in the deployment of learning healthcare models and applications, and the use of a wide variety of data sources and modern methods. Help realize CU-AMC as a national leader in learning healthcare implementation.
- **Build talent.** Aid campus in attracting and retaining talent in informatics, specifically in new technologies. Devise improved criteria for recognition of data science research contributors and their contributions in institutional incentive structures. Promote training in informatics for all career levels and members of the research workforce.
- **Metrics.** Create metrics to evaluate use, utility, and impact of CU-AMC data and analytical resources and activities in support of continuous improvement.
- **Diversity, inclusion, and empowerment.** Help guide CU-AMC in efforts to address diversity and equity issues in all areas of research, healthcare, education, and community engagement by designing data driven strategies and programs that themselves promote diversity.

Preliminary Implementation considerations

The following section details specific areas of focus proposed for the 12 month launch phase of the CRIO office. This phase will start with detailed requirements gathering, benchmarking, and partnership discussions, followed by strategic planning processes to fully identify stakeholder needs and

institutional opportunities that are best served or supported by the CRIo office; it will shape institutional strategy and policies to accelerate innovative, data-intensive collaborative research across the campus. A 5 year CRIo plan would be generated at this time and would be iterated upon and approved by (*Insert leadership group here, presumably VC Research, Dean SoM, ideally Hospital CIOs, etc.*). To regularly keep leadership apprised, the CRIo office will create a mission-wide dashboard for insight into the health of the institution regarding data and analytics. The CRIo office's role is to provide leadership and vision in a catalyst role for numerous stakeholders; those identified to date are listed in the Appendix.

Partnerships and governance.

A foundational and primary role of the CRIo office in service to all campus and affiliated stakeholders.

Goals: Determine and support new partnerships, governance structures and processes, and regulatory and security compliance status and goals for all campus/hospital stakeholders over data and analytical resources. The CRIo office will create a "CRIo council" of key campus and partner stakeholders that will, in collaboration with the CRIo office staff and CU-AMC community, determine requirements, opportunities, risks, and evaluation metrics for campus-wide informatics and data resource development and deployment present and future. The CRIo office and council will consist of a diverse team and will use data to advance equity at CU-AMC. The CRIo office will assist campus partners in making larger purchases and/or design pilots and define evaluation criteria for decision making. The CRIo also recognizes that patients are partners at CU-AMC, and as such, patient engagement, trust, regulatory compliance, and community collaboration must also be taken into account.

Outcomes: Creation of good governance will enhance partnerships among all campus stakeholders; this has tangible scientific, care, and financial impacts. Metrics will be created to support iterative evaluation and prioritization of new advancements. Measures of scientific impact include assessment of improvements in scholarly outcomes, such as publications, collaboration metrics, and the diversity and extent of the funding portfolio. Measures of care relate to the broader use of data and analytical assets to support learning healthcare and other healthcare operational advancements and partnerships. Shared campus measures of financial success include not only the funding portfolio, but also metrics relating to public-private partnerships, commercialization, and related activities realized in collaboration with CU Innovations.

Risks if not implemented: Poor engagement across campus has already led to duplication of effort, stakeholder needs being unmet, and lost opportunities. A lack of further action may result in a waste of resources, potential for data misuse and regulatory or ethical violations, a mistrust among partners internally and externally, and lost revenue from public-private partnerships.

Create a CU Data Ecosystem: strategically leverage data and analytical assets to maximize impact.

Creating a data and analytics nexus will increase research capacity and impact. You can't use it if you don't know you have it.

Goals: Design and support a vibrant and collaborative CU Data Ecosystem. This will include the creation of a strategy for discovery and sharing of campus data and analytical assets by the CU-AMC community. Well-delineated access permissions and discovery mechanisms in combination with robust

attribution and provenance can create and sustain commercialization opportunities. Support and incentives will be provided for improved data management strategies, such as metadata authoring, versioning, data publishing, and data harmonization. The CU Data Ecosystem will include a data index, with data stores in multiple places depending on user needs; analytical framework(s)/compute infrastructure with the ability to access data based on the the data index; support for methods/analytics development; secure and regulatory-compliant infrastructure for data management; training for all types of users/career levels; and results and data publishing and archiving for transparency and reproducibility. The architecture will be coordinated and deployed in collaboration with campus partners (see Appendix) and will include cloud based/on-premises data management, CU-AMC campus authorization access control, security, and regulatory compliance. The CRIo will develop standard operating procedures and standardized performance metrics for participating campus units and shared resources. Partnering units/resources will receive additional FTE to aid requirements determination, analytical advancement, and researcher support. The Library and Research Office will have the shared task of assisting in dissemination.

Outcomes: Improved communication about and coordination of data resources across campus can be measured by usage and contribution statistics, as well as degree of interoperability and/or connection (e.g., analytical platforms or research connecting and using the data assets). Pilots to integrate existing resources and platforms will support capacity building and evolution of governance and infrastructure design. More and complementary investments around data and analytics yield more usage and return on investment. More and stronger program project/collaborative grant applications in a data-intensive world. Improvement in marketing and partnerships with appropriate external audiences will also help identify new commercialization, partnership, and development opportunities. The fraction of mid-scale compute hardware shared across departments will increase. Such metrics will be defined by the CRIo office in collaboration with the community, and will drive a culture of evaluation and continuous improvements.

Risks if not implemented: Inability to attract and retain key informatics faculty and staff. Systems not built fit for purpose waste money and time and can at worst lead to poor science, regulatory compliance, or security violations. Also, storage is expensive, particularly cloud-based storage, and as we move more and more towards hybrid and/or cloud hosting, duplication has not just time costs but recurring financial costs to the institution. Finally, there are lost opportunities for leveraging expertise and time for innovations and collaboration, rather than spending time on basic infrastructure and maintenance aspects of our work.

Identify requirements and coordinate analytics deployment and advancement

Make it easier to do the right thing, faster and better. Lift everyone up to take best advantage of CU-AMC assets and opportunities.

Goals: The primary role of the CRIo's office for analytics is to identify needs and coordinate deployments that rely on integration between data, expertise, and platforms to perform analyses; as well as to promote analytical outcomes being used within the context of learning healthcare and research innovation. Partnerships with existing campus analytics leaders (see Appendix) will help define and realize a multi-modal, secure analytics environment that meets a wide variety of stakeholder needs. Such an environment (perhaps a further evolution of EUREKA) will connect to campus data resources, Health Data Compass, and other data resources. The hardware integration and harmonization efforts and maturing environment will support sharing and reuse of analytical resources,

attribution, provenance, and maximization of compute resources. Data science and analytical training coordination for all career levels and campus stakeholders will enable research, collaboration, and will expedite learning healthcare and translational informatics. Investigators will be supported in their research applications analytics feasibility to ensure robust environment and methods in CU-AMC grant applications. An EHR analytics governance and implementation program will be established to coordinate key initiatives across campus.

Outcomes: The CRI office will be able to document improved analytic capacity in terms of people participating in analytics; degree of hardware and software integration; data resources utilized in the analytics; research innovations materialized by the analytics; new revenue streams and public-private partnerships centered on data and analytics; learning healthcare models deployed by the analytics; funding obtained due to innovative analyses; and resource allocation metrics. Such advancements will also improve the national recognition of CU-AMC as a leader with improved rankings, more alignment of research with health system needs that can generate cost savings, and support of a cycle of reinvestment into research for improved care.

Risks if not implemented: Poor return on investment for expensive and important resources; e.g., a paucity of users, content and quality are not well designed for maximal use/reuse by the community, data is not well managed and therefore cannot serve the multiplicity of users and needs (e.g., learning healthcare, translational research, basic research, public health, etc.). Potential for a flattening grants success rate due to not being competitive in an analytical environment or capabilities.

Foster new commercial partnerships and innovation centered on data and analytics

Aid commercialization of CU-AMC informatics-related technologies and advances.

Goals: Data is an institutional asset that is required for advancing the research and care missions of the institution. Making data more accessible and robust will facilitate these missions, and also provides the opportunity for commercialization and new or enhanced public-private partnerships. Commercialization will also provide sustainability, promote creative solutions and new technologies for licensing, and support incentive structures that value good data practice. The goal is to identify novel informatics technologies, services, or data assets developed at CU-AMC, so that, in partnership with CU Innovations and through pilot funding and workshops, the CRI office can promote commercialization opportunities across the CU-AMC research enterprise.

Outcomes: The CRI office will identify, coordinate, and track informatics technologies and data assets suitable for commercialization opportunities. Metrics will include the number of new commercial partnerships and new investigators, start-ups, and licensed technologies as well as the number of extramurally funded projects using Health Data Compass and other campus resources. Metrics will be devised in collaboration with CU Innovations and the research office.

Risks if not implemented: Data is more valuable the more purposes it serves. A lack of investment in commercialization would reduce the incentives and sustainability of building a robust data ecosystem at CU-AMC. A lack of commercial partnerships also would diminish the reputation of CU-AMC as a national leader in data analytics and informatics.

Enable risk-aware security and compliance processes for research and data management

Balance robust security and regulatory compliance with access and analytical capacity. Make the most effective research approaches the compliant ones by default.

Goals: All clinical institutions must manage risk of data breach. The goal is to carefully balance and mitigate risk with advancements that come from provisioning data access and analytical efficiencies. There is no right answer, only one that maximizes care outcomes and research impact, reduces burden and costs, and supports robust patient privacy. The CPIO office will coordinate key stakeholders, obtain expert opinion, document and help implement decisions, and monitor the outcomes of those security and regulatory decisions in an ongoing fashion.

Outcomes:

Coordinated management, communication, and evolution of security and regulatory requirements will result in a more secure and more efficient research environment. Innovations in data access provisioning can support auditing and provenance for research reliability and attribution. Patients will be satisfied about how their data are contributing to new discoveries and better outcomes, while simultaneously provisioning less risk for the institution of legal or reputational harm. Measures of success will include traditional metrics such as increased publications and awards using increasingly sophisticated access/data types, but also will include documentation and publication of advancements that demonstrate new modes of secure, responsible, ethical data sharing. The CPIO office will also work with campus technology leaders to create metrics for iteratively evaluating security and regulatory compliance.

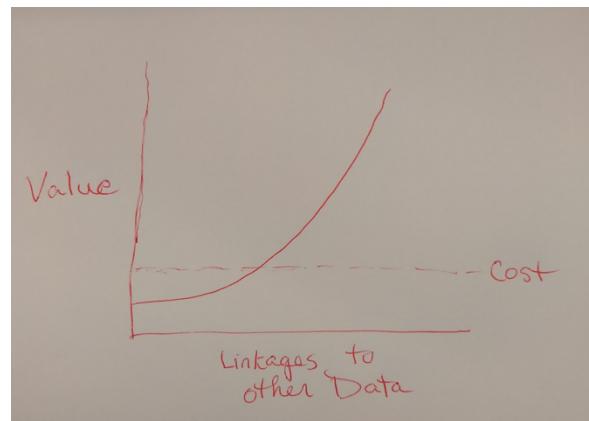
Risks if not implemented: The risks of not having good security and regulatory policies, a robust security system, and widespread understanding and compliance are enormous. Conversely, many organizations err on the side of excessive security rules and infrastructure, reducing research capacity, efficiencies, and actually leading to greater risk due to investigators working around the policies and infrastructure.

Complementarity with existing efforts and resources

CU-AMC has demonstrated commitment to campus-wide investment in informatics research, training, implementation, and infrastructure. The Center for Health AI (and an upcoming biomedical informatics department) will support faculty hiring, research, and training in informatics; the Colorado Center for Precision Medicine focuses on biobanking and serving UC Health and public-private partnerships; the Colorado Clinical and Translational Sciences Institute (CCTS) supports clinical and translational research using EHR and trial data; The Center for Integrated Data Analytics (CIDA) supports researchers informatics and statistics needs; and the Cancer Center Shared Resources support data storage and 'omics data analysis. All of these and other campus endeavors require robust and secure access to EHR data through Health Data Compass as well as integration with other data and analysis tools. The CPIO will perform the necessary socio-technical engineering and coordination processes to maximize institutional investments in the aforementioned resources and realize these within the proposed CU Data Ecosystem and governance structures.

Integration capacity

Many of the roles, responsibilities, and investment funds described below are focused on building informatics and infrastructure capacity by integrating and advancing existing resources. Data is increasingly valuable to research, clinical care, and innovation and commercialization the more linkages it has to other data. Return on investment in generating, cleaning, and maintaining data also is greater upon integration. Further, integration of infrastructure can reduce burden and costs and improve overall efficiency of the informatics ecosystem. We therefore envision the CROI office having a large focus on integration capacity.



Appendix: CU-AMC Stakeholders identified to date

Constituents and partners span the entire range of research discovery from basic biological sciences, to patients, to clinical care, to national data networks. Listed here is an incomplete set of stakeholders based on conversations to date, and focused regionally.

Informatics units/initiatives:

- The Division of Biomedical Informatics and Personalized Medicine (DOM)
- The Section of Informatics and Data Science (DOP)
- The D2V Dean's Transformational Grant Award
- The new Center for Health AI
- The Strauss Health Sciences Library

Informatics resources:

- The Cancer Center Shared Resources
- The Department of Biostatistics and Informatics and the Center for Integrated Data Analytics (CIDA) (CSPH)
- Colorado Center for Personalized Medicine (CCPM), which encompasses Health Data Compass, TICR, and the Personalized Medicine Biobank
- The CCTSI Informatics Core
- The Adult & Child Consortium for Health Outcomes Research & Delivery Science (ACCORDS)
- New Center for Health AI
- RNA Biosciences Initiative

Schools/departments/offices

- All four schools
- Basic science departments
- CU Innovations
- IRB
- University IT/Security

Training:

- Computational Bioscience graduate training program
- Biostatistics and Informatics Department (CSPH)
- Bioengineering Department (College of EDC)

Partners external to CU-AMC:

- Children's Hospital Colorado (CHCO)
- UCHealth
- VA Hospital
- Other campuses: Boulder and CSU
- Patients and community