Rady Faculty of Health Sciences

Complex Data Strategy

2018

Rady Faculty of Health Sciences | University of Manitoba


Rady Faculty of
Health Sciences
University
of Manitoba


HEALTH
inCOMMON


GEORGE & FAY YEE
Centre for Healthcare Innovation


Research
Manitoba

Table of Contents

Background

To provide direction (with respect to complex data) over the next 5 years, the Rady Faculty of Health Sciences engaged in a process that included a stakeholder survey, a full-day workshop (see Appendix 1 for workshop agenda) and a facilitated planning session (see Appendix 2 for session notes). The process was guided by a committee with representatives from the Rady Faculty of Health Sciences (see Appendix 3 for list of committee members).

Prior to a facilitated planning session a survey was distributed to stakeholders, resulting in 24 completed surveys (response rate of 48 percent). Data gathered in the survey provided direction to planning session participants regarding strengths, challenges, opportunities and risks in the current environment (Appendices 4, 5 and 6). Key points from the survey are summarized below:

Strengths

Across the three areas of interest (data infrastructure, data management, and data analytics), a notable strength was quality. Many respondents have access to high quality databases (e.g., comprehensive, large sample sizes, etc.), highly trained staff in data analytics and management, and have high quality data management processes in place.

Challenges

Limited resources were commonly cited challenges. Many respondents indicated limited financial resources to maintain and update/upgrade existing data infrastructure and management and analytic processes. Recruiting additional experienced staff and limited administrative and IT supports were also noted across all three areas as common challenges.

Risks

The lack of a long-term plan for maintaining and upgrading data infrastructure, management, and analytic processes was seen to threaten the long-term sustainability of existing data. Data quality can be compromised if it becomes outdated or is incomplete. Respondents thought investments in the most recent tools and methodologies are required to achieve the best results and remain competitive in the academic environment.

Opportunities

Opportunities for multi-disciplinary, cross-sector collaborations in terms of research and the sharing of resources (databases and infrastructure, IT support, administrative supports, etc.) were noted most commonly. In addition, respondents indicated opportunities exist for the standardization of processes and policies surrounding data infrastructure, management, and analytics (e.g., best practices). Greater data linkages could be established and additional training opportunities are necessary to ensure qualified, experienced individuals are available to support the use of complex data.

Attended by approximately 35 stakeholders, the planning session provided an opportunity for participants to consider the desired impact of the strategy and potential actions that would address need related to data infrastructure and data analytics.

Based on the information gathered during the planning session, a draft framework, including value propositions, strategy purpose, goals and objectives, was developed (see Appendix 7 for a visual disgram). In order to measure progress on identified goals and objectives, a quarterly reporting template, including outputs and outcome indicators, was developed (see Appendix 8).

"The lack of a vision and strategic plan for maintaining and upgrading data infrastructure is a risk to potential long-term sustainability." (Survey Respondent)

Framework for Planning

Purpose

To translate complex data into usable knowledge

Value Propositions

Patient
To improve quality of care
To ensure confidentiality and security of information

Faculty
To enhance data quality and access
To strengthen effective knowledge translation with the public, government and university
To streamline processes and minimize duplication
To support knowledge synthesis
To create an environment that enables state of the art research

Government and Regional Health Authorities

To enhance data quality and access
To inform health and public policy decision making
To inform health policy and care
To optimize resource utilization

Industry
To enhance competitive advantage preemptive

Goals and Objectives

Goals and objectives (for the next 5 years) provide the framework for annual planning, including the development of outputs and short-term outcomes (with baselines and targets) for quarterly reporting.

1. Strengthen alignment and engagement

a. Increase internal alignment within Faculty
b. Increase involvement with the Provincial Information Management & Analytics (PIMA) strategy, Research Improvement Through Harmonization in Manitoba (RITHiM) and University of Manitoba (U of M) central administration
c. Increase collaboration with other faculties and academic institutions
d. Increase inclusion of external stakeholders in governance
e. Strengthen provision of consultation services

2. Enhance data quality and access through cleaning, coding, classifying and retrieval

a. Increase understanding of current capacity relative to best practices
b. Improve recruitment and retention expertise
c. Increase and improve training for highly qualified expertise
d. Improve consistency of high quality data
e. Enhance drive to discovery and innovation

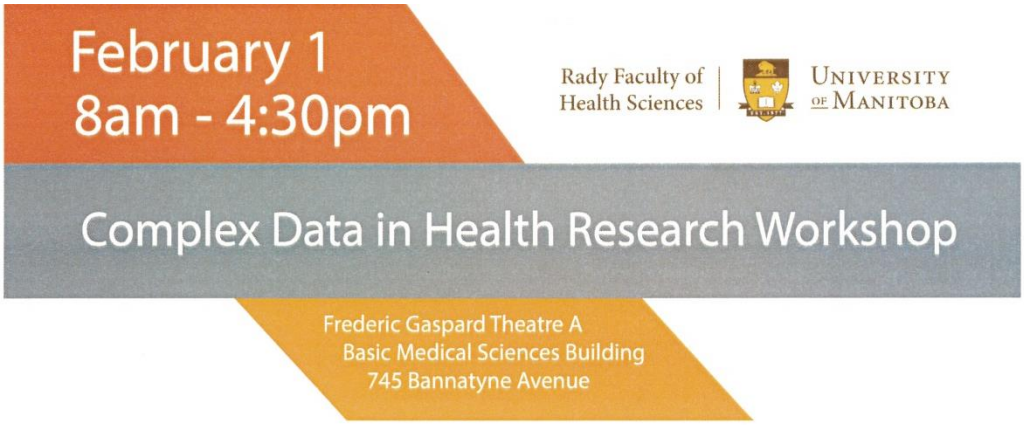
3. Enhance ability to describe, visualize and model complex data

a. Increase understanding of current capacity relative to best practices
b. Strengthen recruitment and retention of highly qualified expertise
c. Increase and improve training for highly qualified expertise
d. Increase utilization of consistent analytical tools
e. Enhance drive to discovery and innovation

Next Steps

- Formulate Implementation Team.
- Share Complex Data Strategic Plan with stakeholders who attended the facilitated planning day session.
- Implement action plan activities and measure progress using the quarterly reporting template (and predetermined outputs and outcome indicators).

Appendix 1: Complex Data in Health Research Workshop Agenda



Theme	Schedule	Agenda Item
	8:00-8:15am	Introduction - ***Peter Nickerson*** - *Distinguished Professor of Internal Medicine and Immunology and the Vice-Dean Research, Rady Faculty of Health Sciences at the University of Manitoba and* ***Lisa Lix*** - *Professor, Department of Community Health Sciences & Director, Data Science Platform, George & Fay Yee Centre for Healthcare Innovation*
System Level Architecture	8:15-8:35am	Provincial Analytics ***Deborah Malazdrewicz*** - *Executive Director, Information Management & Analytics, MHSAL*
	8:35-8:55am	Big Data Strategies for Research and Innovation ***Tim Murphy*** - *Vice President, Provincial Platforms & Alberta Strategy for Patient-Oriented Research (SPOR) SUPPORT Unit*
	8:55-9:15am	Integrating Clinical Data Systems to Demonstrate Improvements in the Performance, Reach, and Value of Manitoba's Healthcare System ***Ryan Zarychanski*** - *Assistant Professor, Department of Internal Medicine, Sections of Hematology/Medical Oncology and Critical Care, University of Manitoba*
	9:15-9:45am	Panel Discussion - *led by* ***Peter Nickerson*** - *Distinguished Professor of Internal Medicine and Immunology and the Vice-Dean Research, Rady Faculty of Health Sciences at the University of Manitoba*
	9:45-10:00am	Break
Data Infrastructure	10:00-10:20am	Data Infrastructure to Support Health Research Innovation in Manitoba ***Lisa Lix*** - *Professor, Department of Community Health Sciences & Director, Data Science Platform, George & Fay Yee Centre for Healthcare Innovation*
	10:20-10:40am	21st Century Data Strategy ***Paul Terry*** - *Chief Innovation Officer, Personalized H-e-alth Management Initiative (PHEMI)*
	10:40-11:30am	Panel Discussion - *led by* ***Peter Nickerson*** - *Distinguished Professor of Internal Medicine and Immunology and the Vice-Dean Research, Rady Faculty of Health Sciences at the University of Manitoba*
	11:30-1:00pm	Lunch and Poster Presentations
Data Management	1:00-1:20pm	Managing the Data Warehouse - 25 Years of Experience at MCHP ***Mark Smith*** - *Associate Director, Repository and Deliverables, MCHP*
	1:20-1:40pm	Test Driving the Data: The Importance of Knowing its Strengths and Limitations ***Malcolm Doupe*** - *Associate Professor in the Department of Community Health Sciences, Faculty of Health Sciences, College of Medicine, and a Senior Research Scientist with the MCHP*
	1:40-2:00pm	Managing Big(ger) Data from The Manitoba Personalized Lifestyle Research (TMPLR) Study ***Peter Jones*** - *Canada Research Chair in Functional Foods and Nutrition*
	2:00-2:30pm	Panel Discussion - *led by* ***Charles Burchill****, Associate Director Data Access and Use, Manitoba Centre for Health Policy*
	2:30-2:45pm	Break
Data Analytics	2:45-3:05pm	Operational Analytics: Providing Accurate and Timely Information to WRHA Decision-Makers ***Trevor Strome*** - *Director, Analytics and Business Intelligence*
	3:05-3:25pm	Genomics in the Age of Big Data ***Gary Van Domselaar*** - *Chief, Bioinformatics, National Microbiology Laboratory – Public Health Agency of Canada*
	3:25-3:45pm	What Skills are Needed in the Analysis of Complex Data in Health Research? ***Rob Balshaw*** - *Senior Biostatistician, George & Fay Yee Centre for Healthcare Innovation*
	3:45-4:15pm	Panel Discussion - *led by* ***Lisa Lix*** - *Professor, Department of Community Health Sciences & Director, Data Science Platform, George & Fay Yee Centre for Healthcare Innovation*
	4:15-4:30pm	Closing – ***Peter Nickerson*** - Distinguished Professor of Internal Medicine and Immunology and the Vice-Dean Research, Rady Faculty of Health Sciences at the University of Manitoba

Appendix 2: Facilitated Session Notes

Big Questions

Participants were asked for their 'big' question going into the day.

- Governance as obstacle
- How to connect evidence to action
- How/where to fit unstructured data
- How to recruit good data scientists
- How to partner multi-faceted research groups
- How to engage/work with indigenous partners
- How to expand rehab data
- Timelines of project in terms of business
- How to integrate info streams
    - Data from different disciplines, platforms, sets
- Grapple with privacy (in timely way)
- Where to invest?
    - Priorities, needs
- How to use radiology data
- Timely, accessible clinical data use for practical use
- Ensure quality data
    - Info management: critical investment for quality data

What is the Desired Impact?

In groups of four, participants drafted and shared a vision/impact statement for discussion in the large group.

- Vision needs to include:
    - Access
    - Usable resources
    - Which is connected and secure
    - To benefit the community (MB and global)
- High quality research that drives policy and care
    - Support health systems and research without creating barriers
- Support the ability to answer questions:
    - People, infrastructure, and relationships
- Develop a system for streamlined and timely integration of data and linkages from different sources that meets the needs of all stakeholders
- Excellence in training and research in data sciences that will advance knowledge and transform health
- Ethical integration/efficiency/effectiveness and capacity in data science to serve our stakeholder/gov't/RHA/faculty/industry
- Creating a critical mass of multidisc. Experts to provide competitive service

Based on initial ideas for a shared vison, Health in Common and the large group drafted and discussed options for a final statement.

(draft 1)

Rady Faculty drives policy and care through integration and excellence in research and training.

Rady Faculty [~~drives~~ informs health policy and care] through integration and ~~excellence~~ innovation in research and training.

Knowledge generation

- complex data research and training
- data management
- broad sense of health determinants

Improves the health of Manitobans decision making in public policy and health care

(draft 2)

The Rady Faculty generates knowledge through complex data research and training to enhance health and public policy decision making

In Partnership?

Generates change / transformation, tools, resources

Data as a commodity

Development?

(draft 3)

The Rady Faculty improves health outcomes, quality of care, and resource utilization by using and translating complex data into knowledge. through research and training.

Identifying Actions for Focus Areas

Data Infrastructure

Participants were asked to individually identify what needs to happen with respect to the focus area. Next, in small groups, participants shared and clustered similar actions.

Group 1

- Fit for purpose

- Usable data
- Facilitate data usage

- Maintaining "research ready" data is costly! (but important!)
- Need to organize to identify players for various collaborations
- Better infrastructure (people, technology, methods) to support integration
- Guidelines/policies to facilitate integration of variable data sets

- Data quality
- Necessary time to increase data quality needs to remain as important
- High quality data – linkable
- New approaches (people, technology, methods) to enhance data quality

- Skilled analysts and data miners
- Human resources, multi-disc. competencies linked to training
- Trained and experienced users

Group 2

- Data accessibility
- Greater accessibility/integration
- Be able to access data from anywhere
- Accessible data like of all relevant data
- Accessible linked data

- Increased capacity building at all levels
- More scientists and trainees with expertise in area
- Bioinformatic support to provide data analysis to researchers
- Better communication around what's available
- Centralization and integration of province wide databases
- Adoption of common standardized formats and search functions
- High quality data
- Secure open access to "data library" managed by data curator
- Database managers, IT infrastructure

Group 3

<u>Create/optimize new infrastructure</u>

- Plan infrastructure with MB Health within larger university and national context
- Create metadata standards
- Transition to cloud
    - $ resource management, ~~$ servers $~~

<u>Training/recruitment</u>

- Ensure data quality
- Data curation
    - Quality
    - Metadata
- Data infrastructure
    - Improve data curation and quality
- Recruit HQP in data curation

- Proper data management
- Accessibility

    - Well supported process to access the data
    - Human resource ongoing training of researchers
- Accessibility
- Provide efficient and responsible access
- Ethical access
    - Privacy, OCAP, security, timelines
- Access is critical
    - Uses mechanisms which give access based on data needed, defined privileges of research

Group 4

- Accessible, primary as well as secondary data user inspired infrastructure inspired by users
- EPR (electronic patient records) and research
    - Supports access
    - Privacy barrier

- Ability to align diverse data
    - E.g. genomics and bio-imaging to extract meaning
- Data map/ Architecture design
    - How organized
    - Who/how administered
    - How maintained
    - Who responsible / ownership
    - Who is authorized access
        - segmented
    - Data types
- Take stock of data we want to focus on (ROI)
- Capacity
    - What data
    - What data types
    - How much data/how long
    - [impacts on research]

- Recruitment of best and brightest
    - How to recruit
    - How to retain
- HR availability to run infrastructure – training or funding
- Standardize positions within AESES for analysis?
    - How to define [classify? Description?] class
        - Re: competition with industry / other provinces'
- Increase technology awareness
    - What is available to support data, access, use – marketing of what's available
- Secure and sustainable infrastructure that is modifiable/flexible enough to grow with future need and to address diverse stakeholders

- How to get data out for data in? Demonstrate our value
    - Real time clinical data

Data Infrastructure Opportunities

In small groups, given the identified actions, participants identified data infrastructure opportunities.

Group 1

- Better integrated data
- Data collection quality control
- "fit for purpose" now and in future
    - Nimble – how to collect based on new tech?
- Are we collecting the right type of data?
- "sufficiently fluid"
- System to system communication
- Multi-disciplinary human resources
- Guidelines for collecting data to facilitate integration
- Training re: data entry
- Communication feedback

Group 2

- Accessibility
- Increased collaboration
    - Agreements between institutions share data
- MB Health is looking at PIMA
    - Opportunity for the university to partner. We're part of the discussion
- Shared Health Services development
    - Opportunity for greater accessibility as changes are made (eHealth data)
- Changing technologies/computer power – take advantage
- Shared storage
- Engage computer scientists with expertise to develop innovative practices

Group 3

- RITHIM: Active participation in implementing
- Explore tools and opportunities around secure access
- Address risk averse environment among data stewards
    - Implement systems, review standards
- Integrated distributed data infrastructure [principle, multi partners]
- *Train and recruit / establish team of skilled data scientists to 'curate' data so that it's research ready*
- *Support continuous improvement of front-line data collection*
- *Education and KT to increase awareness and understanding importance of good data*
- *Build on existing strengths re: data linkages*

Group 4

- Share information to enhance awareness to identify gaps
    - Take stock (“metadata data”)
    - Potential use (where you want to go)
    - Integrated planning (with informed people)
- Recruitment of “talent” – who is there?
    - Use data as recruiting tool (multiple across disciplines)
    - Leverage MB advantage (small but big)
- Use identifiable data to improve patient care
    - Be aware of boundaries viz. MB Health
    - RTC
- Privacy by design (build it info infrastructure)
    - Opp to share good “pockets” of knowledge
    - Supports accessible data
- ROI update foundational data sets
- Data management – utility focused infrastructure (in terms of closed projects and eliminate detractors), address duplication

Data Infrastructure Priorities

Finally, each small group identified and reported back 3 priorities related to data infrastructure.

Group 1

1. Better systems to enhance data quality
2. Data and expert inventory: what's out there and what's possible?
3. Design standards to better integrate:
    - Linkage
    - Coding
    - Hardware
    - Governance
    - Consistency

Group 2

1. Engage stakeholders to develop a data library (centralized), managed by a data curator
2. Create training programs to support the development and use of data library
3. Support "proof of concept" projects of data library (demonstrate the infrastructure is working, linkages, how)

Group 3*** (is this correct?)

1. Train and recruit / establish team of skilled data scientists to 'curate' data so that it's research ready
2. Support continuous improvement of front-line data collection
3. Education and KT to increase awareness and understanding importance of good data

Group 4

1. I.D. talent and share talent/knowledge
2. Take stock of gaps (data mapping) to ensure we know what our investments (ROI, utility focused)
3. Privacy by Design
    - I.D. what is well designed
    - Celebrate what we are good at

Further discussion in large group to begin identifying common priority areas:

- Data library with curator (development and use)
- Design standards
- Inventory of expertise, including frontline → develop plan for training and recruitment
- Take stock of gaps – nuts and bolts
- [privacy by design] – privacy and ethical standards
- Increase understanding and awareness of NB of good data

Data Analytics

Participants were asked to individually identify what needs to happen with respect to the focus area. Next, in small groups, participants shared and clustered similar actions.

Group 1

- Access to sufficient infrastructure (connected, accessible technology, speed, capacity)
- Develop integrated programs to support expanding role of data analytics across disciplines
- Improve training in analytics
- Access to tools, supports, function specific open, access data often, access to training and support
- Ensure data analytics tools are available to all that need them
- Adoption of open source free software as the go-to standard for analytics (transferrable skills)
- Don't reinvent – there are many "black box" analytic tools
- We need to look forward to AI as an analytic tool
- Improve BI – improve user friendliness
- Define, elaborate, understand analytics versus analysis versus decision making
- Sharing of resources
- Create a core group of experts to provide support in data analytics related to complex data
- Increase training program

Group 2

- Resource sharing
- Increased offering of more complex programming workshops (e.g. SKS/R) if they exist already, increase awareness
- Training, at many levels
- Develop analytic capacity – training, recruitment, mentorship
- HR resources to support analytics – training or finding, what area? What is the right mix? Multiple individuals in the "right" teams
- Coordinated, skilled HR training
- Balancing rigor and programmatic
- Need great data visualization, build capacity
- Develop the best system to be used for analytics
- Involvement and pre-designed analysis plans at the design/data collection phase
- Platform for analytics – general analyses (SAS. R, etc.), general utilities (python, etc.), visualization, big data – correlation, causation, association
- State of the art up to date hardware, software, analytical tools

Group 3

- Support for ongoing training and development in analytics to remain current and cutting edge across Rady FHS
- Ability to access data and support and data handling
- Training on how to {illegible} data and systems with central data entry

- Capacity to focus on {illegible} and limitations of data, focus analytical capacity on gaps
- Data processing and tools, analytic skills training, processing tools to normalize data, categories of data analysis
- Integrating different types of analyses by a single individual
- Centralized capacity versus researcher's capacity – clarity needed
- Measure and track where current resources are going and develop HR plan
- Increased availability and access to statistical and bio-informatics expertise across Rady FHS
- Balance of in-house development and "off the shelf" – investment, ROI
- Develop and recruit bio-informatics, novel methods
- Short courses, workshops needed to keep up to date in new advances in analytics

Group 4

- Unify and integrate data
- Communication i.e. different stakeholders need to bring people together to increase scope of discussion to capture different expertise
- Generate meaningful insights
- Training program in spectrum analytics
- Need for ongoing training
- Train the trainers – those that can share their expertise and experience to improve the base level of analytics globally
- Address gaps in analytics capacity and expertise – bioinformatics, computer biology
- Complex data = complex analytics (rapid change in analytics is challenging to keep up with)
- Academic silos making progress challenging
- Multi-disciplinary integration and collaboration of bioinformatics
- Coordination and sharing of data analytic workflows and pipelines for maximizing return on investments and timelines
- Depth of expertise required in how the data are generated – e.g. which sequencing platform used
- Contributes to MB's economic engine

Data Analytics Opportunities

In small groups, given the identified actions, participants identified opportunities related to data analytics.

Group 1

- Design training programs to develop the workforce to support data analytics in industry, academic, government (all stakeholders)
- Identify, develop, and evaluate tools
- Create group of analytic experts
- Knowledge and inventory of existing resources and personnel
- Breakdown barriers between faculties and colleges
- No silos between users and analysts
- Collaborative research – with cities, institutions, etc. to exchange expertise

Group 2

- Resource sharing (faculties and beyond) – everyone does not need to be an expert in everything
- Follow the steps of other provinces regarding sharing resources and expertise (e.g. Alberta)
- Harnessing the expertise in MB
- Provincial level changes – being aware and changing with them, being a part of it
- Taking advantage of training opportunities

Group 3

- Develop, recruit and retain
- Understanding clarify analytic skills sets needed
- Increased expertise of Rady faculty, access and availability, where to go to access data analysis support – map it
- Better understand the research question
- Training – statisticians, clinicians, state of art approach
- Process to measure what model is built – no duplication
- Data documentation – meta data
- Documentation of programming – sharing methodology

Group 4

- Cording and access – coordinate access of data analytic workflows and pipelines for maximizing return on investments and timeliness
- Tools for data analysis
- Know what we have – unify and integrate the data, communication (bring people together to increase scope of discussion to capture different expertise)
- Training – ongoing, spectrum analysis, train the trainers

Data Analytics Priorities

Finally, each small group identified and reported back 3 priorities related to data analytics.

Group 1

1. Training – design programs to develop workforce, supporting data analytics for all stakeholders
2. Tools – identify, evaluate, develop
3. Teamwork – no silos between users and analysts, breakdown barriers between faculties and colleges, collaborative research to exchange expertise (cities, institutions, etc.)

Group 2

1. Gap and overlap analysis (SWOT), including HT training and tools
2. Comprehensive and strategic training and ensuring we have the expertise to provide the training
3. Increased collaboration, including analysts in earlier stages, sharing resources, co-creation, etc.

Group 3

1. Multi-disciplinary data sciences institute at the U of M
2. Major recruitment initiative in data science – shared positions between university and stakeholders – research practitioners
3. Integrate service level analytics

Group 4

1. Map analytic skill sets and needs, including general and specific expertise
2. Training for specific and general audiences – methods and tools
3. Process to evaluate emerging tool (ROI)

Further discussion in large group to begin identifying area

- Data library with curator (development and use)
- Design standards
- Inventory of expertise, including frontline → develop plan for training and recruitment
- Take stock of gaps – nuts and bolts
- [privacy by design] – privacy and ethical standards
- Increase understanding and awareness of NB of good data

Considering Values and/or Principles

A large group brainstorm identified initial thoughts regarding values and/or principles.

- Include people providing data
- Access and efficiency
- Capacity
- Training
- Innovation
- Challenging health of Manitobans
- Quality
- Integration
- Learning healthcare system
- Impact beyond border
- Answer complex questions
- Discovery research
- Benefit stakeholders
    - Faculty
    - Communities – RHA, gov't, industry, Indigenous Manitobans
    - Patient
- Impactful
- Intersectoral
- Learning environment
- Intelligence generating
- Secure and accessible
- Partnership

- Innovation – packaging things differently
- Consider Return on Investment – don't try to integrate everything
- PRINCIPLES – to what end?
    - Patient, public, community outcomes
    - change how we deliver care
    - essential partner
    - get more, higher quality research to provide support to the health system

Appendix 3: Committee Members

The steering committee members for the Complex Data Strategic Plan were:

- **Peter Nickerson**, Distinguished Professor of Internal Medicine and Immunology and the Vice-Dean, Research, Rady Faculty of Health Sciences
- **Charles Burchill**, Associate Director, Data Access and Use, Manitoba Centre for Health Policy
- **James Davie**, Professor, Department of Biochemistry and Medical Genetics; Scientist, Children's Hospital Research Institute of Manitoba; Senior Scientist, Research Institute in Oncology and Hematology; and Director, Manitoba Epigenetic Network
- **Frank Krupka**, Executive Director, George & Fay Yee Centre for Healthcare Innovation
- **Lisa Lix**, Professor, Department of Community Health Sciences and Director, Data Science Platform, George & Fay Yee Centre for Healthcare Innovation
- **Christina Weise**, Chief Executive Officer, Research Manitoba

Appendix 4: Data Analytics

Strengths

Human Resources, for example:

- Highly trained, experienced analysts/programmers with specific knowledge to enable effective analysis of data (e.g., SAS, clinical epidemiology, biostatistics, database analysis, pharmaceutical, etc.)
- Ability to do all the data analysis personally from alignment to interpretation using Perl, Python and R for analyzing, visualization and statistics; writing your own shell scripts to make the pipe lines

Processes and capabilities, for example:

- NVivo analysis capacity
- Well-documented
- Organized
- Systematic
- Automated
- Repeatable
- Cluster computing for improved efficiency

Challenges

Quality issues, for example:

- LIS systems

Training issues, for example:

- Training of staff and students in existing and new data analytic methods is a potential challenge (staff/students have difficulty accessing and using data properly on their own; need increased training on SAS)
- Lengthy process to fully train
- Increased familiarity with macros available to process variables/conduct analyses

Lack of coordinated data management despite experts/bioinformatics experts present, for example:

- numerous tools and techniques available but limited streamlining and integration

Limited Resources, for example:

- supports in data organization
- Lack of space to keep the raw and processed files and time delays due to limited server capacity
- need upgrades to hardware and software to stay current and cutting edge
- Expansion into other software packages, computer support
- Labour intensive

Opportunities

Multidisciplinary teams/collaboration and consolidation of supports, for example:

- bioinformatics experts in different departments should be consolidated under one leadership and vision
- bioinformatics support group, interfacing with CHI biostatisticians
- sharing of resources

Increased educational opportunities, for example:

- E.g., educating physicians on the importance of this process and providing training to help them understand the analytic processes to support collaborative work
- The biostatisticians in the Department of Community Health Sciences can serve as faculty-wide resources for advanced analytics of complex health data
- Increased SAS training
- Workshops for faculty and students

Resources and methods, for example:

- local servers with restricted access in the department
- machine learning and FTA, visualization
- improve user-friendliness of data access

Risks

Resources, for example:

- Lack of human resources (e.g., limited qualified staff, overworked staff due to increasing requests for grants/projects; limited number of individuals to teach/train)
    - Need to outsource
    - Could incorporate skillset into undergrad and postgrad medical educations
    - Risk that courses in advanced analytics, which are offered by CHS may be oversubscribed in terms of the number of students; can negatively impact student learning experience
    - Potential for errors increases
- Wasted resources and expertise due to inefficient processes
- High costs associated with server maintenance
- Costs associated with implementing the latest techniques

Complexity, for example:

- Complexity of analysis deters new users

Appendix 5: Data Management

Strengths

Well-defined processes, for example:
- data management and support processes
- storage (e.g. local hard drives)
- access within a secure environment
- training
- Validation and quality control at ingestion, documentation, approval process
- Clear understanding of OCAP, and how to negotiate data ownership and use

Breadth and quality of data, for example:
- Various oral health information
- linkages to different types of information
- metadata
- Scrambled PHIN based linkage
- New LIS will solve some technical issues (linkage, data quality, standardization within DSM)

Challenges

Consistency and compatibility, for example:
- Some databases are not provincial
- Not centralized or consistent within units/organizations
- Diagnostic databases not interfaced with clinical databases
- Comparisons between old and new data (e.g., historical lab-generated vs. contemporary lab-generated data)
- Transport of data and linkages suboptimal
- Differences in data management processes – creates duplication in terms of effort
- Data not consistently backed up and processes used not consistent/efficient
- Documentation, ingestion process means consistent installation configuration, human resources.

Training, for example:
- researchers now need to articulate data management plan in grant applications

Data management, for example:
- Individualized
- Minimizing risk of re-identification
- Retrieval of past users' information difficult (e.g., passwords from prior students)

Limited Resources, for example:
- Expensive, intensive requirements to keep track of permissions
- Ongoing maintenance
- De-identification of free text and supporting metadata is difficult due to fiscal challenges
- MCHP limitations - only one RAS, occasionally slow

- Sufficient number of qualified staff interested in data management processes

Opportunities

Improvements to accessibility & use, for example:

- Better interfaces/linkages with other databases
- data collection at additional healthcare facilities
- More independence and better computing
- Ability to export methodology into a standalone package

Best Practices/Standardization, for example:

- harmonization of data management with expert support; expertise in handling big data
- Centralized process to keep track of permissions, share process with other data providers; automating PHIN scrambling without reliance on MHSAL
- Policy development and proper training
- Expansion of use and linkage opportunities, sharing of Validation/Quality review control processes

Risks

Limited resources, for example:

- suboptimal technical environment and limited IT support
- lack of human resources
- lack of coordinated efforts between research and data management
- Potential for duplication of effort
- reliance on MHSAL to scramble PHINs for linkage

Storage, for example:

- currently no overarching management of storage capacity and big data; need long-term planning and flexibility
- Loss of data due to damage to physical environment (fire, flood, etc.)

Maintenance, for example:

- Time consuming to keep up to date with changing data management processes
- Metadata are at risk of being out of date

Access, for example:

- Increased demand can increase the complexity of the data access process and increase time it takes to get access
- Data security
- Not being the sole owners of data that is housed within the centre/organization

Appendix 6: Database Infrastructure

Strengths

Accessibility, for example:

- Online
- Central server/ multiple servers
- well supported processes to access the data with appropriate permission

Comprehensive, Population based data, for example:

- diverse array of social and health databases and niche databases owing to our unique subject population/circumstances (e.g., Manitoba Centre for Health Policy)
- databases for international studies (e.g., Centre for Global Public Health)
- substantial longitudinal data (e.g., Manitoba Follow-up Study)
- High quality (e.g., multiple years of data; large sample sizes; multiple measurements)
- De-identified but linkable

Data, methodology, and systems, for example:

- metadata
- Updated equipment and systems that are secure and backup regularly
- Simple processes – such as using regular computers and RAS to MCHP

Human Resources, for example:

- trained individuals; analytic experience of staff; opportunities for collaboration
- interest from students to learn
- ongoing training of researchers Re: OCAP and other similar frameworks
- strong IT team

Challenges

Access and Compatibility, for example:

- Multiple different storage medias
- Non-open proprietary software formats (e.g., ACCESS, LIMS, etc.)
- Delays in being able to access necessary data (approval, costs, etc.)
- Documentation of data
- Obtaining novel datasets and linkages, similar to what exists in other provinces
- Current electronic databases not user-friendly
- potential for re-identification possible (this can also be seen as an advantage)
- ability to access data from only limited designated physical location(s)
- limited access nodes when multiple researchers/students need access

Lack of consistent policies across databases, for example:

- OCAP principles must be recognized and adhered to when using data from Indigenous partners for research purposes
- No Department-wide guidance on data storage and management
- Validity of data fields

Limited resources, for example:

- Capacity
- Limited knowledge and experience in selection and use of databases
- Limited administrative and IT supports
- maintaining "research ready" data and metadata is costly
- Limited clinical data
- data infrastructure and databases require frequent updates to ensure they meet the requirements for cutting-edge data storage and analysis

Opportunities

Centralized data & coordination of database management and supports, for example:

- Provincial database with provincial support system for public and private sectors; interfaced with clinical databases and areas beyond health to add greater context and impact on health overall
- A central server to secure data better; provide password protected access; wider access with VPN
- new hardware/software and methodologies (e.g., NLP and machine learning; automated visualization; CANARY connection)
- addressing under-utilized databases (e.g., lab, post-secondary student data)
- key individuals or super-users to support critical functions and shared administrative support; collaborative training

Multidisciplinary collaborations, for example:

- Cross-jurisdictional opportunities (e.g., internal and external partners)
- Increased opportunities to use data for thesis projects
- Collaborations within RFHS and other U of M departments/faculties (e.g., Biostatisticians within CHS to address methodology challenges associated with complex data working with researchers skilled in knowledge translation strategies for analyses of the data)

Expanded IT/analyst supports, for example:

- Analysts from other disciplines (e.g., computer science)
- Increased computational capacity with bioinformatics experience; greater support for Mac systems

Risks

Accessibility & management, for example:

- Not accessible to all faculty and students (e.g., due to access issues, costs, etc.)
- Consent issues and improper use of databases
- Slow to adapt to changing demands
- Multidisciplinary/cross-jurisdictional work can create issues with approvals and resources (financial/personnel)

Limited resources, for example:

- Degradation of metadata due lack of updates for budgetary reasons or lack of time invested
- new analytic techniques are very resource intensive requiring new hardware and software

- limited IT supports
- Lack of good websites/information for the institutions and departments to attract and recruit the strongest candidates (faculty, students, staff)

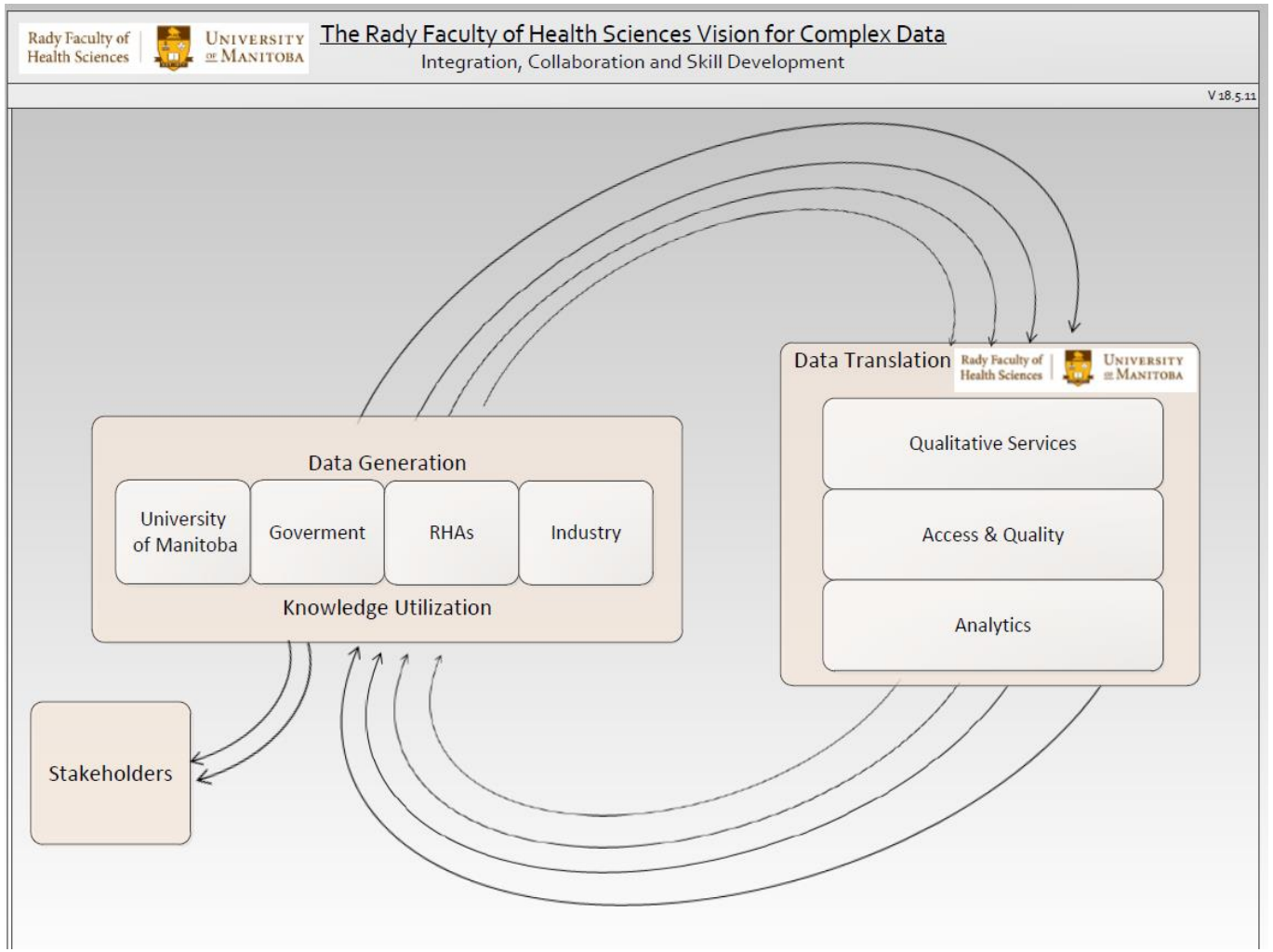
Potential for re-identification/ security, for example:
- Public awareness and misunderstanding of identification risk

Single source, for example:
- Loss of data
- relying solely on one model (e.g., MCHP)

Data quality, for example:
- Necessary time to improve data quality is not being invested
- Incomplete fields; inconsistent data input formats
- Fragmentation
- Data migration issues from old systems to new with inherent risks and potential loss of data
- Not maintaining the most current analytic tools
- Rapid increases in data sizes/types will exceed current capacity (can inhibit research)
- Open access can threaten quality of data

"The lack of a vision and strategic plan for maintaining and upgrading data infrastructure is a risk to potential long-term sustainability." (Respondent)

Appendix 7: The Rady Faculty of Health Sciences Vision for Complex Data Diagram



Appendix 8: Complex Data Quarterly Reporting Template

Goal 1: Strengthen alignment and engagement

Objective and Activities **1a. Increase internal alignment within Faculty** i. Create and convene leadership table ii. Identify and implement opportunities for shared learning (e.g. annual big data day, symposia) iii. Create and convene education/training committee		
OUTPUTS	**YEAR TO DATE (** indicates last quarter)**	**COMMENTS**
i. Leadership table meetings held		
ii. Opportunities for shared learning identified		
iii. Training committee meetings held		
OUTCOME INDICATOR (2-3 years)	**BASELINE (Target 2018/17)**	**STATUS**
2/3 of training committee members indicate satisfaction with efforts to increase alignment		
Increased # of shared learning opportunities planned		

1b. Increase involvement with PIMA, RITHiM and U of M central administration

i. Participate on steering committees with U of M Central Admin, PIMA and RITHiM
ii. Ongoing engagement with PIMA and RITHiM working groups

OUTPUTS	YEAR TO DATE (** indicates last quarter)	COMMENTS
i. Steering Committee participation		
ii. Working Groups participation		

OUTCOME INDICATOR (2-3 years)	BASELINE (Target 2017/18)	STATUS
% if identified committees and working groups with regular participation		
% of representatives indicating satisfaction with level of engagement		

1.c. Increase collaboration with other faculties and academic institutions

i. Identify and implement opportunities to co-fund initiatives
ii. Develop and implement common training programs
iii. Participate in and lead regional and national initiatives (e.g. SPOR, PRHDN)

OUTPUTS	YEAR TO DATE (** indicates last quarter)	COMMENTS
i. opportunities to co-fund identified and implemented		
ii. training programs developed and implemented		
iii. Regional initiatives participation and leadership		
iv. National initiatives participation and leadership		

OUTCOME INDICATOR (2-3 yrs)	BASELINE (Target 2017/18)	STATUS
# of co-funded initiatives		
$ value of co-funded initiatives		
# of joint training sessions provided		

1d. Increase inclusion of external stakeholders in governance

i. Engage U of M central table, PIMA, RITHiM, BAM, and Innovative Medicines Canada on Senior Leadership

OUTPUTS	YEAR TO DATE (** indicates last quarter)	COMMENTS
i. Representatives from each stakeholder (U of M central table, PIMA, RITHiM, BAM, and Innovative Medicines Canada) identified and confirmed		
ii. # of Senior Leadership meetings attended by each stakeholder		

OUTCOME INDICATOR (2-3 yrs)	BASELINE (Target 2017/18)	STATUS
% of external stakeholders attending meetings	n/a (100%)	
% of stakeholders indicating satisfaction with level of engagement	n/a (100%)	

1.e. Strengthen provision of consultation services

i. Develop a centralized process for intake
ii. Harmonize existing service delivery
iii. Improve computational biology and omics consultation services
iv. Engage library services

OUTPUTS	YEAR TO DATE (** indicates last quarter)	COMMENTS
i. Centralized process developed and implemented		
ii. Service delivery harmonization opportunities identified and implemented		

OUTCOME INDICATOR (2-3 yrs)	BASELINE (Target 2017/18)	STATUS
# of consultations provided		
% of clients satisfied with services		
% of clients reporting increased capacity post-services		

Goal 2: Enhance data quality and access through cleaning, coding, classifying and retrieval

Objective and Activities

2.a. Increase understanding of current capacity relative to best practices

i. Develop and utilize expertise matrix (i.e. Inventory of people resources)
ii. Identify benchmarking and best practices through systematic review via RFP process
iii.
iv. Engage with regional and national initiatives

OUTPUTS	YEAR TO DATE (** indicates last quarter)	COMMENTS
i. Matrix developed and utilized		
ii. Benchmarks and best practices identified		
iii. Regional initiative participation		
iv. National initiative participation		

OUTCOME INDICATOR (2-3 years)	BASELINE (Target 2017/18)	STATUS
Increased participation in national and regional initiatives		
Increased understanding of current capacity		
Increased alignment with benchmarks and best practices		

2.b. Improve recruitment and retention of expertise

i. Identify and recruit a min. of 3 academic positions
ii. Identify and recruit a min. of 3 technical positions
iii. Identify and utilize funding opportunities
iv. Provide ongoing support for career development
v. Optimize resources to create state of the art environment

OUTPUTS	YEAR TO DATE (** indicates last quarter)	COMMENTS
i. Academic positions recruited		
ii. Technical positions recruited		
iii. Funding opportunities secured		

OUTCOME INDICATOR (2-3 yrs)	BASELINE (Target 2017/18)	STATUS
>50% of technical positions have been with Faculty over 5 years		
Increase # of positions with complex data expertise		
Increased satisfaction with career development, supportive and creative environment and career development opportunities		

2.c. Increase and strengthen training for highly qualified expertise

i. Inventory of current U of M training programs to identify gaps and overlaps
ii. Identify and utilize funding opportunities (e.g. NSERC Create)
iii. Explore and expand opportunities for collaboration with other student training programs
iv. Create interdisciplinary internships
v. Develop, promote and support data science student association

OUTPUTS	YEAR TO DATE (** indicates last quarter)	COMMENTS
i. Gaps and overlaps identified		
ii. Amount of funding identified and utilized		
iii. Internship menu of opportunities developed		
iv. Student Association created and supported		

OUTCOME INDICATOR (2-3 yrs)	BASELINE (Target 2017/18)	STATUS
Increase # of students from other academic programs working with Faculty		
Increase # of Faculty students with external placements		
Increase # of students completing internships		
Level of satisfaction with student association		

2.d. Improve consistency of high quality data

i. Create and utilize ongoing feedback loop on quality of data provided (with data providers)
ii. Develop and utilize pre-emptive strategies to ensure data quality
iii. Develop and provide data quality training for data providers and analyzers
iv. Develop and implement complex data training for executive/supervisory level positions

OUTPUTS	YEAR TO DATE (** indicates last quarter)	COMMENTS
i. Feedback loop developed and utilized		
ii. Training developed and provided		
iii. # of pre-consultation projects		
iv. # of data protocol publications		

OUTCOME INDICATOR (2-3 yrs)	BASELINE (Target 2017/18)	STATUS
Increased # of consultant contacts		
Increased % of consultations that are pre-consults (e.g. grant development)		
Increased satisfaction with data quality		

2.e. Enhance drive to discovery and innovation

i. Provide seed funding (e.g. Rady Innovation Fund, Research MB Team Awards)

ii. Recruit faculty with protected research time

iii. Create funding mechanism for salary support of academic positions (e.g. research Chairs, CRCs)

iv. Increase supports for grant funding

OUTPUTS	YEAR TO DATE (** indicates last quarter)	COMMENTS
i. $ value of seed funding provided		
ii. Funding to support Academic positions secured		
iii. Increased support for grant funding identified		

OUTCOME INDICATOR (2-3 yrs)	BASELINE (Target 2017/18)	STATUS
Increase # of research FTE in the system		
Increase in funding for academic positions		
Increase satisfaction with support for grant funding		
Increase # of successful grant applications		

Goal 3: Enhance ability to describe, visualize and model complex data

Objective and Activities

3.a. Increase understanding of current capacity relative to best practice

i. Develop and utilize expertise matrix (i.e. Inventory of people resources)

ii. Identify benchmarking and best practices through systematic review RFP process

iii. Identify and implement opportunities to leverage current resources

iv. Engage with regional and national initiatives

OUTPUTS	YEAR TO DATE (** indicates last quarter)	COMMENTS
i. Matrix developed and utilized		
ii. Benchmarks and best practices identified		
iii. Opportunities to leverage resources implemented		
v. Regional initiative participation		

OUTCOME INDICATOR (2-3 years)	BASELINE (Target 2017/18)	STATUS
Increased participation in national and regional initiatives		
Increased understanding of current capacity		
Increased alignment with benchmarks and best practices		
Increased participation in national and regional initiatives		

3.b. Strengthen recruitment and retention of highly qualified expertise

i. Identify and recruit a min. of 3 academic positions
ii. Identify and recruit a min. of 3 technical positions
iii. Identify and utilize funding opportunities
iv. Provide ongoing support for career development
v. Optimize resources to create state of the art environment

OUTPUTS	YEAR TO DATE (** indicates last quarter)	COMMENTS
i. Academic positions recruited		
ii. Technical positions recruited		
iii. Funding opportunities secured		

OUTCOME INDICATOR (2-3 yrs)	BASELINE (Target 2017/18)	STATUS
>50% of technical positions have been with Faculty over 5 years		
Increase # of positions with complex data expertise		
Increased satisfaction with career development, supportive and creative environment and career development opportunities		

3.c. Increase and improve training for highly qualified expertise

i. Inventory of current U of M training programs to identify gaps and overlaps
ii. Identify and utilize funding opportunities (e.g. NSERC Create)
iii. Explore and expand opportunities for collaboration with other student training programs
iv. Create interdisciplinary internships
v. Develop, promote and support data science student association

OUTPUTS	YEAR TO DATE (** indicates last quarter)	COMMENTS
i. Gaps and overlaps identified		
ii. Amount of funding identified and utilized		
iii. Internship menu developed		
iv. Student Association created and supported		
v. Gaps and overlaps identified		

OUTCOME INDICATOR (2-3 yrs)	BASELINE (Target 2017/18)	STATUS
Increase # of students from other academic programs working with Faculty		
Increase # of Faculty students with external placements		
Increase # of students completing internships		
Level of satisfaction with student association		

3.d. Increase utilization of consistent analytical tools

i. Develop and utilize framework to select and evaluate tools
ii. Establish a tool implementation committee that applies consistent framework
iii. Develop an inventory of approved tools
iv. Identify funding mechanism to support

OUTPUTS	YEAR TO DATE (** indicates last quarter)	COMMENTS
i. Policy and procedures developed to support framework		
ii. Committee established and active		
iii. Inventory of tools completed		
iv. $ value of funding secured		

OUTCOME INDICATOR (2-3 yrs)	BASELINE (Target 2017/18)	STATUS
Decreased use of tools not included in framework		
Increased use of approved tools		
Fluctuation in level of investment needed to maintain tools		

3.e. Enhance drive to discovery and innovation

i. Provide seed funding (e.g. Rady Innovation Fund, Research MB Team Awards)

ii. Recruit faculty with protected research time

iii. Create funding mechanism for salary support of academic positions (e.g. research Chairs, CRCs)

iv. Promote access to grant funding

OUTPUTS	YEAR TO DATE (** indicates last quarter)	COMMENTS
i. $ value of seed funding provided		
ii. Funding to support Academic positions secured		
iii. Increased support for grant funding identified		
iv. $ value of seed funding provided		

OUTCOME INDICATOR (2-3 yrs)	BASELINE (Target 2017/18)	STATUS
Increase # of research FTE in the system		
Increase in funding for academic positions		
Increase satisfaction with support for grant funding		
Increase # of successful grant applications		