

Treat your Data like an Organizational Asset: Understanding the Data Lifecycle

Geoff Zakaib, P.Eng., MBA

Raising the Bar: Evaluation Collective Conference November 22, 2016

Presentation Outline

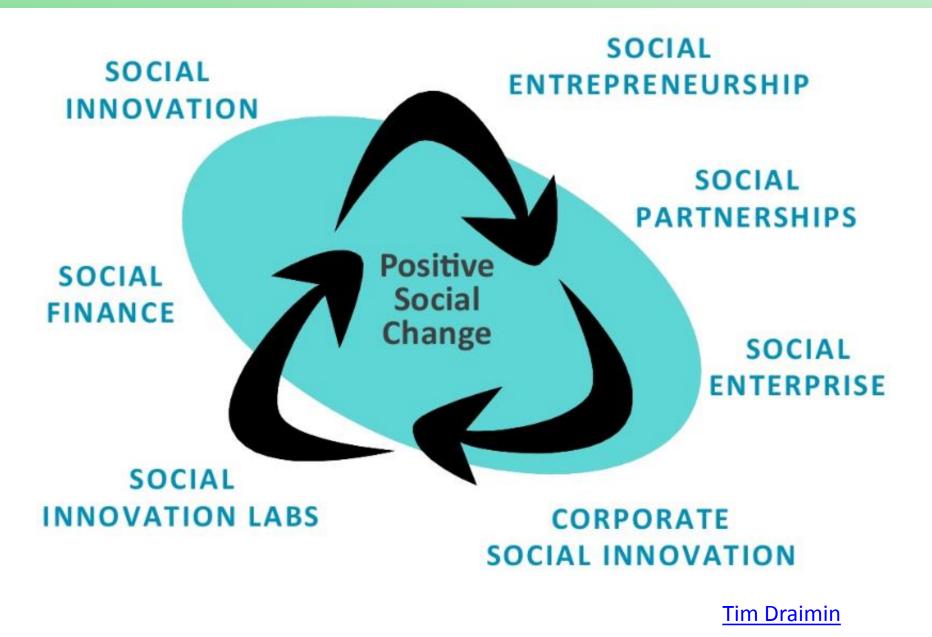
- Social and Technological Context for Evaluation
- Data and Nonprofit Organizations
- Background and Components of the Data Lifecycle
- Nonprofit Technology Trends
- Nonprofit Data Case Studies and Commentary
- Data for Social Good

Nonprofit Mindset Evolution

NEED INADEQUATE BURDEN	\rightarrow	OPPORTUNE STRENGTH CHANGE
SYMPTOMS	\rightarrow	SOLUTIONS
FUNDRAISING	\rightarrow	FINANCING
PROJECTS	\rightarrow	PLATFORMS
NECESSITY	\rightarrow	EMPATHY + LOVE + NECESSITY
INCREMENTAL CHANGE	\rightarrow	SYSTEMS CHANGE
MIN COSTS	\rightarrow	TRUE COSTS

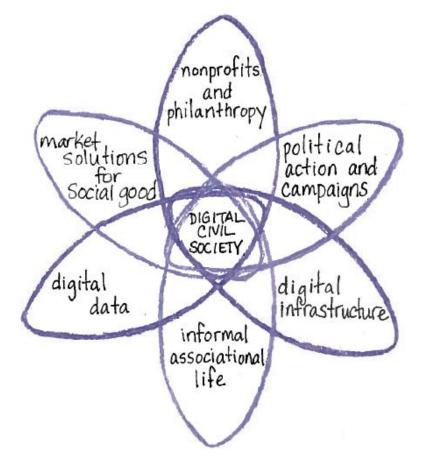
Tim Draimin

Social Ideas and Concepts



Digital Civil Society

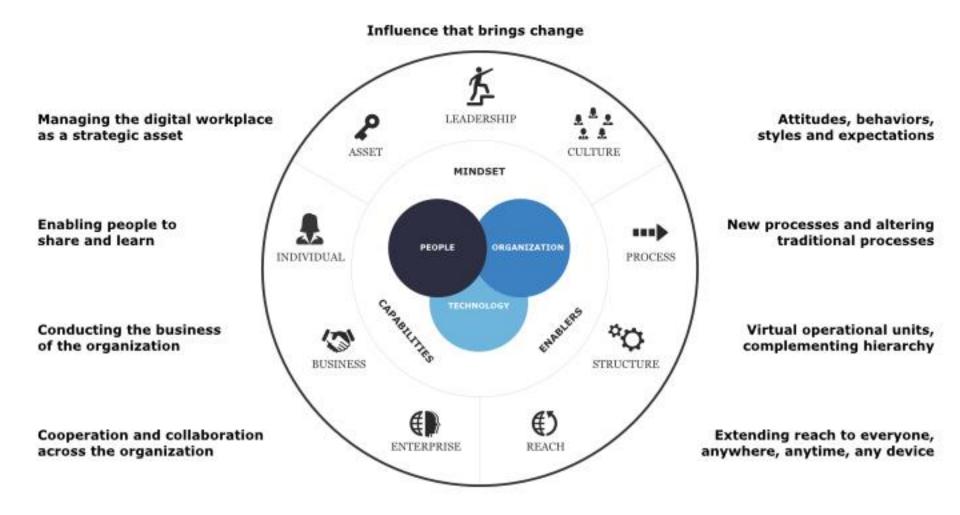
- Civil Society and the Social Economy ... Lucy Bernholz
- <u>Digital Civil Society</u> the ways we use our private resources for public benefit in the digital age



Digital Technology Transformation

- Impact of digital technology
 - Innovation
 - Transparency
 - Collaboration
 - Participation
- Characteristics of digital technology
 - Connective
 - Efficient
 - Intelligent
- Potential to contribute to key challenges by reinventing
 - Public service, often in less costly ways
 - Community, and how people collaborate
 - Business, in ways that are better aligned with human needs

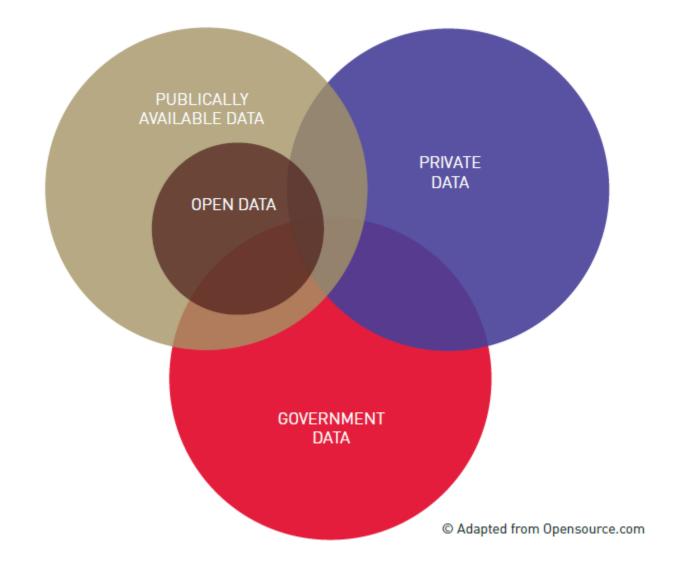
Organizations in Digital Age



Types of Data

- Who We Are Administrative Data
- What We Do Programmatic Data
- Informing Our Work Baseline Data
- Communicating Our Impact Data on What Works

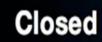
Types of Data





The Data Spectrum

Small / Medium / Big data Personal / Commercial / Government data Internal Named Group-based Public Anyone access access access access Employment Explicitly assigned Via authentication Licence that **Open licence** contract + policies by contract limits use Sales reports **Driving licences** Medical research **Twitter feed Bus timetable**



@ 00

Shared



theodi.org/data-spectrum

Continuum of Information/Data

Personally Identifying Information Used primarily to deliver services

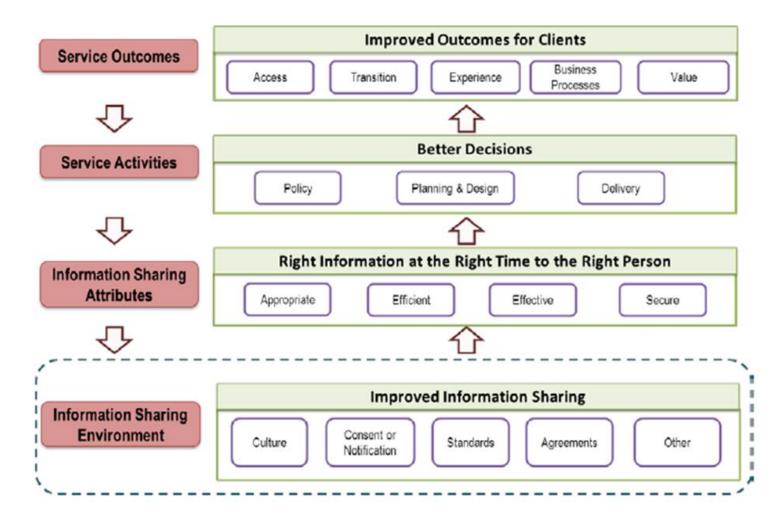
Linked

Data

Administrative Data often linked, used for research and evaluation Open Data Statistical, Non identifying, Population Level, Available for Public Use

Alberta Information Sharing Strategy

Outcomes Drive Information Sharing



Nonprofit Data Strategy

720 Bathurst Street, Suite 405 Toronto, Ontario M5S 2R4

July 2015

Towards a Data Strategy for the Ontario Nonprofit Sector

Data Strategy Key Principles and Components

KEY PRINCIPLES

A few core principles must guide the strategy as a whole:

- Effective use: Nonprofits should put data to effective use to serve their communities not just collect it, but proactively use it.
- Responsible use: Data should be created, collected and accessed responsibly and ethically, with attention to power dynamics that could mar the use of data and with respect to the privacy and safety of those involved.
- For public benefit (not for profit): Nonprofits and governments should be committed and able to access data for public benefit use.

Based on data strategies developed in other areas and on early international and Canadian projects, we have identified four essential components of a successful data strategy:

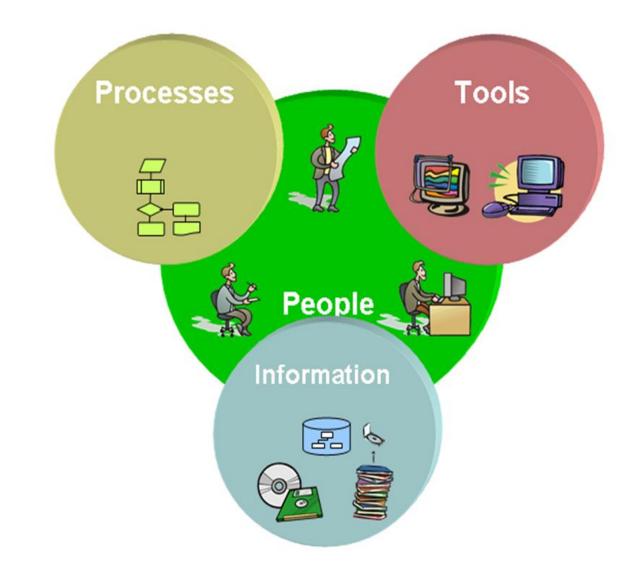
- 1. Standards
- 2. Policy
- 3. Skills & Resources
- 4. Leadership

Data and Nonprofit Organizations

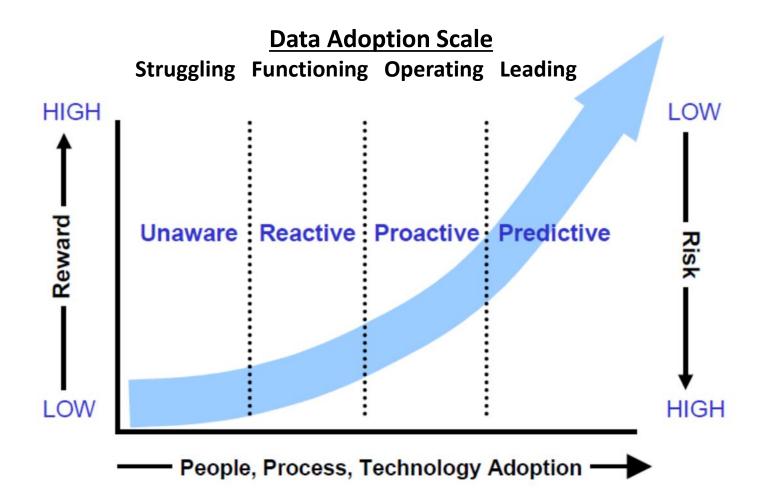
Growing recognition of the need to;

- collect data about programs and operations
- manage the data efficiently and effectively
- make data-informed decisions
- share insights with stakeholders and society

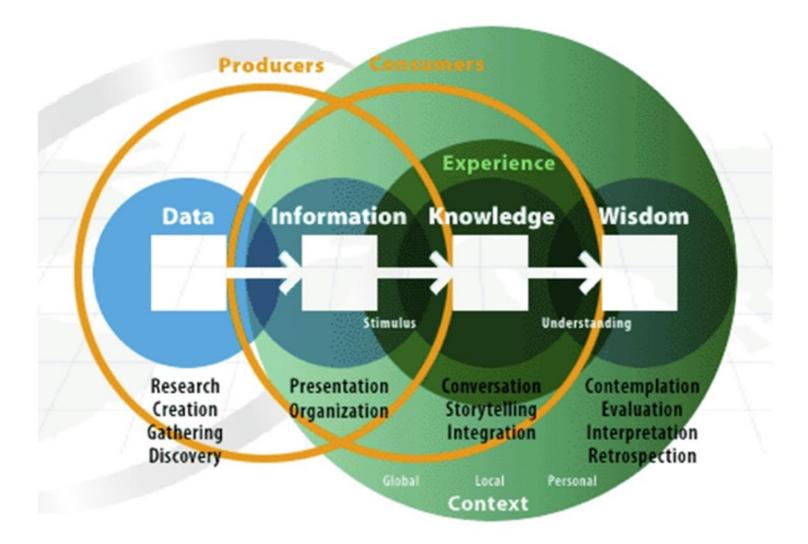
People, Processes and Technology



Data Management Maturity



Data Transformation



Nonprofit Data Management Challenges

Too many systems

• Forces entering the same data into multiple systems with no clear links and no single system of record

Poor data quality

• Makes analysis and reporting difficult and unpleasant

Data are not actionable

• Data are fragmented, yield limited insights, and do not communicate your mission or help you pursue it better

Why is the Data Lifecycle important?

- Data is a valuable asset within nonprofit organizations and must be actively managed – cost of poor Data Mgmt can be very high
- Value of data has to be sustained over the complete life cycle
- Increasing data deluge and risk of data loss can be addressed
- Data can be used to make more effective decisions and to improve operational efficiency.

- Data Lifecycle provides a framework
- Applies to all levels of data management maturity
- People, Processes and Technology
- Data, Information, Knowledge and Wisdom

Data Lifecycle: Models

Data Lifecycle

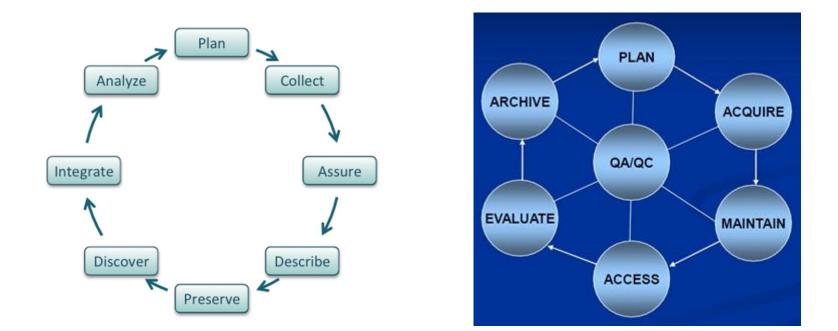
- Reflect the stages from creation to disposal and the various overarching elements – provides consistent terminology
- Incorporates data management principles and practices
- Not always a 'cycle' steps can occur in different sequences or in tandem and some repeated more often than others
- Model adopted should be 'fit for purpose' for organization
- May consider only part of life cycle or combine stages
- Think through flow of data and identify gaps in current practice

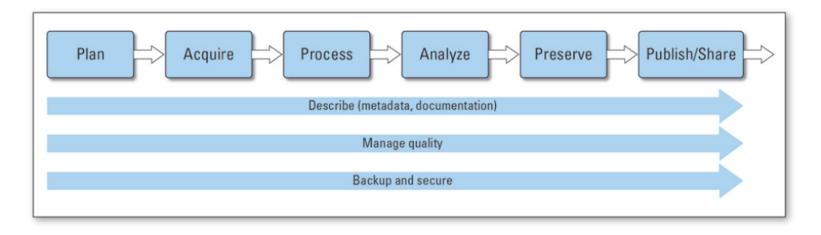
- Context is important data in nonprofit organizations
- Not just technical in fact technology is least important factor

Data Lifecycle: Basic Models

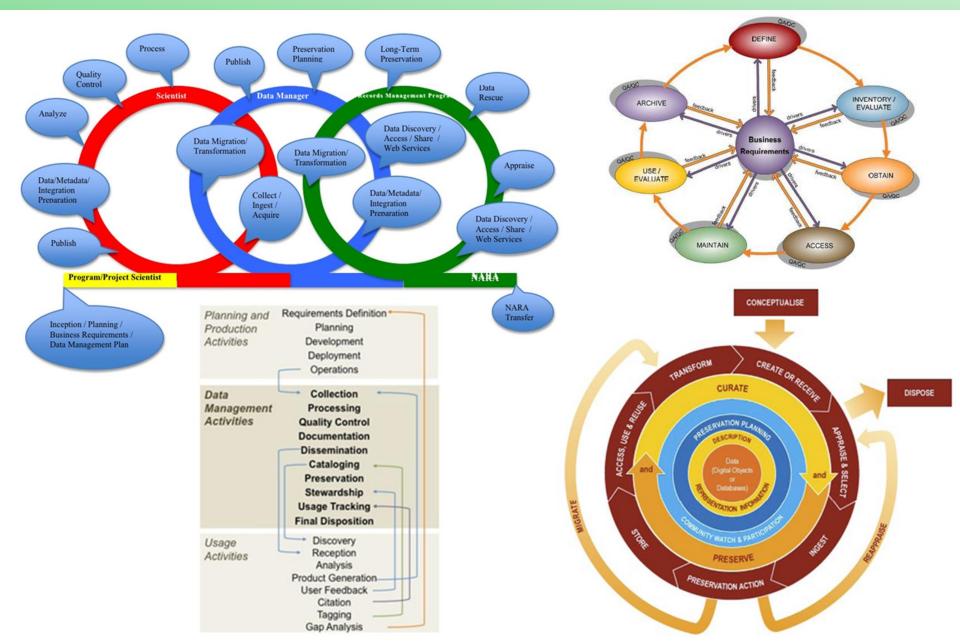


Data Lifecycle: Basic Models

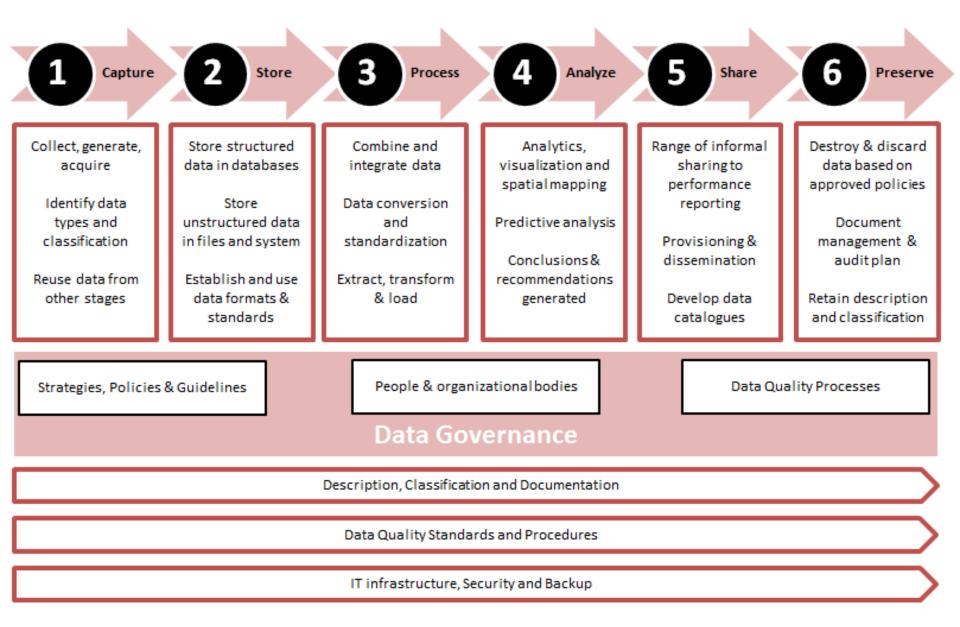




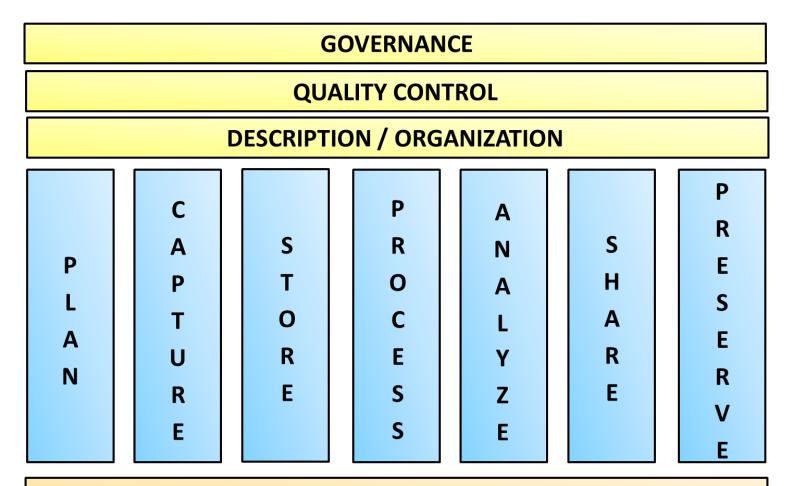
Data Lifecycle: Complex Models



Data Lifecycle Framework



Data Lifecycle Framework



SECURITY / PRIVACY

IT INFRASTRUCTURE

Data Lifecycle: PLAN

Description

• Assess and document data requirements, how the data will be managed and made accessible throughout its life cycle

Additional Aspects

- Strategic and Operational determination of data required
- Data Management Plan, Strategic Information Plan
- User Requirements, Architecture and Technical Design
- Resources Required (people, skills, \$\$'s, technology)

- Value Drivers, Changes over time, Feedback Loop
- Keep it simple and understandable share the plan widely
- Build vs. Buy, Partnering, Funder Considerations
- Government / Regulatory / Society Norms and Expectations

Data Lifecycle: CAPTURE

Description

• Stage that the data becomes available to the organization reflecting the various data types, sources and collection methods

Additional Aspects

- Collect, Generate, Acquire consider Open Data, external sources
- Identify data types for newly collected and existing data
- Re-use data from other stages / life cycles

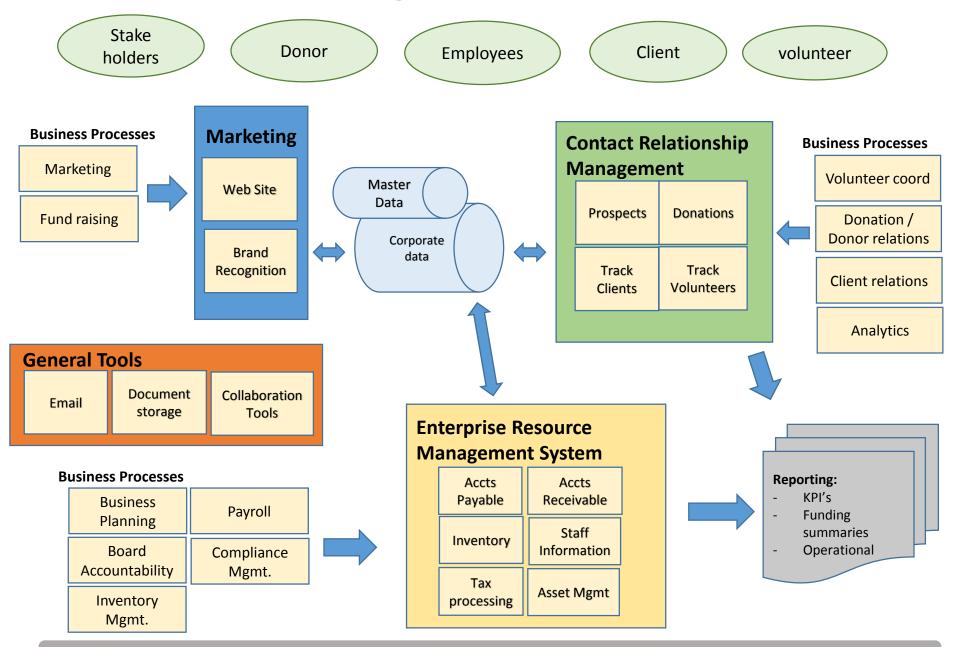
- Critical stage to capture metadata aid to access / discovery
- Data quality is often determined in this stage once it is damaged or destroyed it is difficult (often impossible) to recover

Data Lifecycle: CAPTURE

Types of Data within a Nonprofit

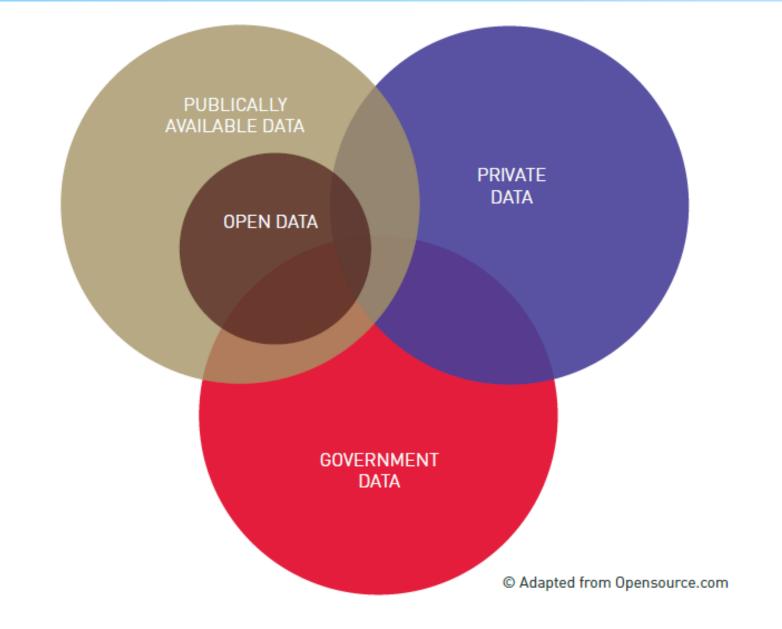
- Administrative Data: Staffing, Development, Volunteer
 Management
- **Financial Data**: Funder Requirements, Donation Management, Grant Management, Budgets, Board Reports, Annual Reports
- **Program Data**: Client Information, Event Attendance, Partner Contact Information, Resource Lists
- External data sources: Referral Lists, Sector Data, Statistics, Economics, Demographics, Social Indicators

General Not for Profit – High Level IT Framework



Infrastructure / Security

Types of Data



Purposes or motivations for doing	Approaches				
something called "evaluation"	Performance Program Measurement Evaluation		Systems Evaluation	Applied Research	
Measurement will be used by an external agent, such as a government department, to determine whether local programs, sites or agencies have delivered a service as planned.	Good fit!	Not a good fit. ≻Evaluation methods are more complex than needed and the	Can work well. ➤When time and energy are invested in shared performance measurement systems.	Not a good fit. ≫Applied research is not designed for day-to-day monitoring.	
Measurement will be used internally by a nonprofit to determine whether individual programs or sites have delivered a service as planned.	≻ If data is used!	turnaround time for analysis may be too slow.	Not a good fit. ➤Shared measurement tools aren't typically flexible or sensitive enough to track nuances of local programming.		
Measurement will be used by an external agent to determine whether local programs, sites, or agencies have achieved impact as planned.	Can work well. ➤If performance measurement systems are sophisticated, specialized, and carefully monitored.	Can work well. ➤ But evaluations undertaken for this purpose may not be as good at generating local insights or actions (see below).	Can work well. ➤When time and energy are invested in shared measurement systems, ongoing communication, backbone infrastructure, and a shared sense of purpose.	Not a good fit. ➤Applied research is not designed to inform action in a direct way.	
Measurement will be used by nonprofits or networks of nonprofits to develop insights about their work and its impact leading them to improve practice.	Rarely works well. ➤Measures are focused on outputs, buy-in is minimal, and analysis is basic.	Good fit! ➤ Especially when time and energy is invested in buy-in, communication, clarity of purpose, and plans for use.	Can work well. ≫If local sites are engaged as partners in the process.		
Measurement will be used by networks representing a sector or the community as a whole to develop new knowledge about best practices and long-term impacts.	Rarely works on its own (although performance measurement methods are often useful when incorporated into more complex	Rarely works. ➤Local evaluations do not typically measure long term change and are not designed to generate generalizable knowledge.	Good fit! ➤Especially when time and energy are invested in shared measurement systems, ongoing communication, backbone infrastructure, and a shared sense of purpose.	Good fit! ➤Especially when the research questions are highly focused and specialized.	
Large systems will use measurement to develop insights about their work and its impact leading them to improve practice.	systems evaluation projects).	Rarely works. ➤It is challenging to aggregate findings from different local evaluations.		Can work well. ➤If research findings are presented in an accessible way and provided in a timely manner.	

Data Lifecycle: STORE

Description

• The methods and technologies that are used to store and persist data within the organization

Additional Aspects

- Structured Data stored in databases (standalone and application)
- Semi-structured and Unstructured Data stored in spreadsheets, documents and content management systems
- Establish and use organizational data formats and standards

- Repository or Data Mart for important structured data
- Document / Content Management Systems for unstructured data
- Backend IT Infrastructure that is fit for purpose

Data Lifecycle: STORE

		Data Format		
		Structured	Unstructured 🎬 📣 💼	
Data Source	Internal	Human-Generated Survey ratings Aptitude testing Machine-Generated Web metrics from Web logs Product purchase from sales Records Process control measures 	Human-Generated Emails, letters, text messages Audio transcripts Customer comments Voicemails Corporate video/communications Pictures, illustrations Employee reviews 	
	Flickr	Human-Generated Number of Retweets, Facebook likes, Google Plus +1s Ratings on Yelp Patient ratings ratings Machine-Generated GPS for tweets Time of tweet/updates/postings 	Human-Generated • Content of social media updates • Comments in online forums • Comments on Yelp • Video reviews • Pinterest images • Surveillance video	

Data Lifecycle: PROCESS

Description

 Data may need to be further processed to put it in a form that is suitable for analysis and sharing

Additional Aspects

- Data from disparate sources may be combined (integrated) to form one homogeneous set of data that can be readily analyzed
- Data conversion / transformation may be required keep raw data
- Extract, Transform and Load (ETL) technology is available

- In this stage there is a high risk of deteriorating data quality
- Versioning control is necessary to track changes and provide audit trail (i.e. who made the change and when)

Data Lifecycle: ANALYZE

Description

• Data are analyzed for insights and to provide a basis for action

Additional Aspects

- Analytics, Visualization and Spatial Analysis (mapping)
- Modelling and Statistics may be used for Predictive Analysis
- Interpretations, Conclusions and Recommendations are generated in this stage

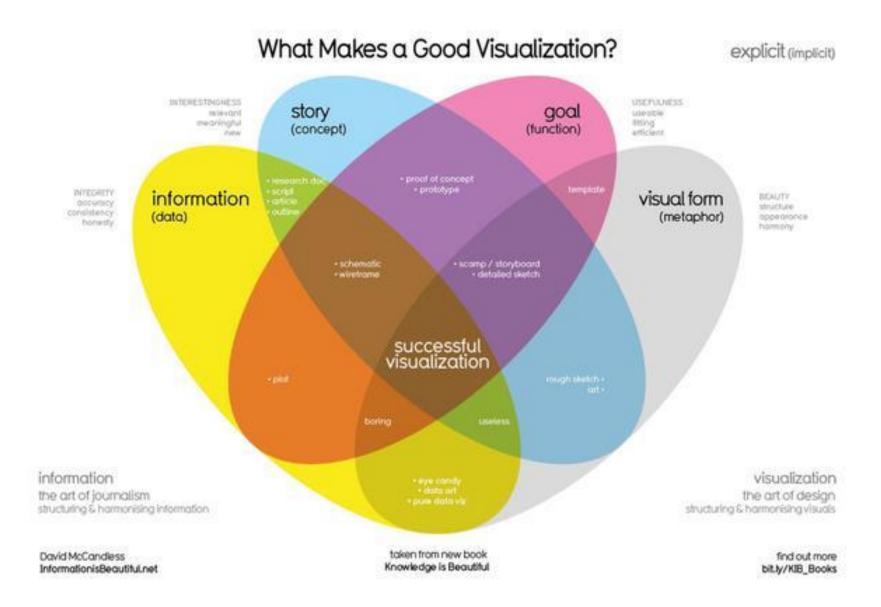
- Specialized Tools, Techniques and Technologies are available
- Basic vs. Complex, Open Source vs. Commercial
- Data Sources and Assumptions must be documented

Data Lifecycle: ANALYZE

4 types of data analytics

Type of analytics	Explanation	Examples				
Descriptive analytics	Provides insight based on past information. What is happening?	Used in standard report generation and in basic spreadsheet functions such as counts, sums, averages, and percent changes and in vertical and horizontal analyses of financial statements.				
Diagnostic analytics	Examines the cause of past results. Why did it happen?	Used in variance analyses and interactive dashboards to examine the causes of past outcomes. Can be used to predict an accounts receivable balance and collection period for each customer and to develop models with indicators that preve control failures.				
Predictive analytics	Assists in understanding the future and provides foresight by identifying patterns in historical data. What will happen? When and why?					
Prescriptive analytics Assists in identifying the best option to choose to achieve the desired outcome through optimization techniques and machine learning. What should we do?		period of accounts receivable and to optimize the				

Data Lifecycle: ANALYZE



Data Lifecycle: SHARE

Description

 Raw, processed and interpreted data are provisioned to others internally and/or externally

Additional Aspects

- Uses range from informal sharing to formal performance reporting
- Dissemination, Publication and Research as end-state 'products'
- Development of Data Catalogues aid discovery and sharing

- Privacy and security are important concerns in this stage
- Front end technologies such as websites, portals and dashboards
- Social media can provide challenges to governance and control
- Confidentiality and Intellectual Property must be respected

Data Lifecycle: PRESERVE

Description

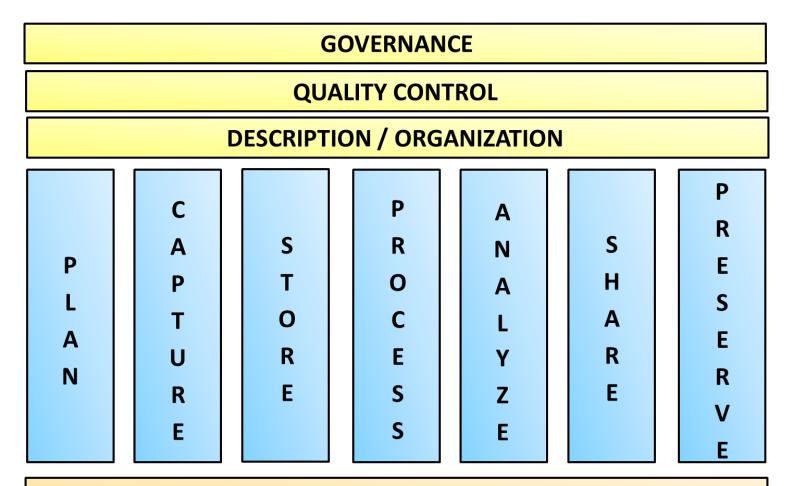
 Sustainment over the long-term by preserving data in an appropriate filing system or archive – What data? For how long?

Additional Aspects

- Discarding / Destruction of some data is legitimate
- Records / Content / Document Management and Audit Plan
- Formats and Standards should be carefully selected

- Preservation of metadata is important
- Various levels of technology sophistication available (i.e. local file system or database, remote data centre, cloud)
- Legal rules and regulations must be considered (retention policy)

Data Lifecycle Framework



SECURITY / PRIVACY

IT INFRASTRUCTURE

Data Lifecycle: GOVERNANCE

Description

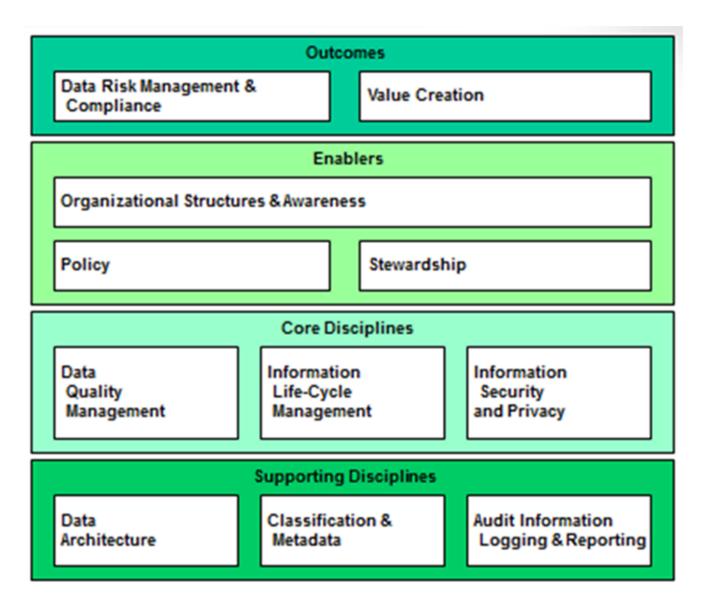
 Roles, policies and procedures used in the exercise of decisionmaking and authority for data-related matters

Additional Aspects

- Data Risk Management
- Data Owners, Data Stewards, Data Producers, Data Consumers
- People and Change Management

- Adoption of data standards
- Training and reinforcement over time
- Management needs to demonstrate support 'walk the talk'

Data Lifecycle: Elements of Effective Governance



Data Lifecycle: QUALITY CONTROL

Description

Data quality is assured across the life cycle through QA/QC processes, monitoring and control

Additional Aspects

- Built-in quality starts at capture and continues across life cycle
- Validation, Quality Monitoring / Assurance and Audit / Inspection
- Master Data Management as a means to identify quality issues and to sustain / improve data quality over time

- Completeness, validity, consistency, timeliness and accuracy
- Identify missing data and/or duplication
- Tools are available for data quality profiling and data cleansing

Data Lifecycle: DESCRIPTION / ORGANIZATION

Description

 Data are accurately and thoroughly described (e.g. who, why, what, when, where) and organized using appropriate standards and organizations principles

Additional Aspects

- Metadata standards are deployed across the life cycle
- Essential for data to be understood, interpreted and used
- Data organized according to approved structures / taxonomies
- Versioning, Indexing, Cataloguing

- Data discovery is a key consideration
- Positive reinforcement needed to combat human nature

Data Lifecycle: SECURITY / PRIVACY

Description

• Systems and processes that ensure data is secure and that both internal and external access is appropriately controlled

Additional Aspects

- Version Control, Backup Procedures, Access Rights and Privileges
- Privacy Protection, Confidentiality, Intellectual Property
- Policies for data access, sharing and re-use

- Government Laws, Regulations and Policies (i.e. PIPA, FOIP, HIA)
- Password protection and multi-form authentication
- Technology must support organizational responsibilities avoid single point of failure (built-in redundancy)

Data Lifecycle: IT INFRASTRUCTURE

Description

• The foundation of technology and services upon which data management activities across the lifecycle are built

Additional Aspects

- PC's, Servers, Databases, Data Backup & Restore, Networks
- Analytic Tools and Applications, Data Sharing Infrastructure
- Ongoing operation, maintenance and support

- Tech Support is an important bridge to user community
- Tech for Non-profits <u>TechSoup</u>, <u>Microsoft</u>, <u>Google</u>
- <u>Nonprofit Tech for Good</u>, <u>Nonprofit Technology Network (NTEN</u>)

Data Lifecycle Resources

- USGS Data Management Website <u>Data Life Cycle Overview</u>
- Data Management Planning Considerations <u>Checklist</u>
- DataONE: <u>resources</u> for data management planning
- DMPTool: <u>free online tool</u> for creating data management plans

Nonprofit Technology Trends

- Social Technology Landscape
 - It's about building communities, fostering new ways of collaborating, and guiding these efforts to achieve a purpose
- Digital Technology Strategy
 - Digital technologies like mobile, social and analytics compliment rather than compete with current systems and information, enabling greater leverage with less disruption
 - Digital strategy brings together digital and physical resources resulting in innovation rather than disruption
 - Concentrate on specific outcomes rather than implementing grand strategies giving the effort a clear and manageable focus
- <u>Digital Transformation</u> Impact on nonprofit value chains
- <u>Digital isn't optional</u> Values Aligned Technology

Digital Technology Assessment

- <u>Steps to Evaluate and Optimize Nonprofit's Use of Digital Tools</u>
- Organizational Assessment of Digital Tools
 - About your organization: General Organizational Info
 - How your Organization communicates
 - How your Organization fundraises
 - Your Organization's Website
 - Your Organization's Email Strategy
 - Your Organization's Social Media Strategy and Use
 - Other online tools
 - How your organization collaborates and backs up data
- <u>Data Maturity: Where Does Your Nonprofit Stand?</u>

Technology Implementation

• Tech Is Everyone's Job

- Make Technology Part of Everyone's Job
- Don't Stop Training on Day 2
- Shared Investment in Outcomes
- Creating Space for Innovation
- Idealware Field Guide to Software for Nonprofits
- <u>Technology Resources for Social Innovators</u>
- Innoweave Practical Tools for Social Innovation

Nonprofit Data Case Studies and Commentary



COLLECTED VOICES DATA-INFORMED NONPROFITS JANUARY 2014







markets FOR

Making Sense of Data and Information in the Social Sector

Five Ways to Avoid the Data System Trap

Using these <u>five strategies</u> will greatly increase the odds that you are developing your performance management data system on a strong foundation:

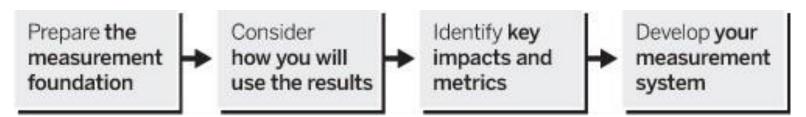
- 1. Don't start the process until your <u>theory of change</u> and indicators are in order.
- 2. Go through at least one data collection cycle to ensure the tools and reports are solid.
- 3. Make sure that everyone who will contribute to the system has a voice in designing it.
- 4. Use a cross-functional planning team to get a crossfunctional system.
- 5. Treat it as a change management, not a technology, project.

Measuring and Improving Social Impacts

The Social Impact Creation Cycle



The Impact Measurement Roadmap



The Role of Data in Program Evaluation

Nonprofits should be focused on producing outcomes and we need evidence that nonprofits are delivering on those outcomes. However, the processes and methods we currently use to do so are in need of improvement: <u>Andrew Means article</u>

- 1. Prove or Improve?
- 2. Periodic or Ongoing?
- 3. Causal Inference or Prediction?
- 4. What's Next?

Data is changing the social sector. We need so much more than what we have today. We need to use new skills, new methods, and new models that embrace data and promote the value it brings to the sector. As that happens our organizations will improve, our effectiveness increase, and we can begin to make greater progress in solving some of our world's most intractable problems.

Better Data About Nonprofit Programs

As nonprofit organizations strive to produce data on performance, effectiveness and impact the biggest headaches are likely to be: <u>Laura Quinn article</u>

- 1. Data quality.
- 2. Providing data to funders.
- 3. Meeting changing data requirements.
- 4. Defining how best to measure for improvement.
- 5. Trying to measure impact.
- 6. Fending off bad research.
- 7. Proving your value.

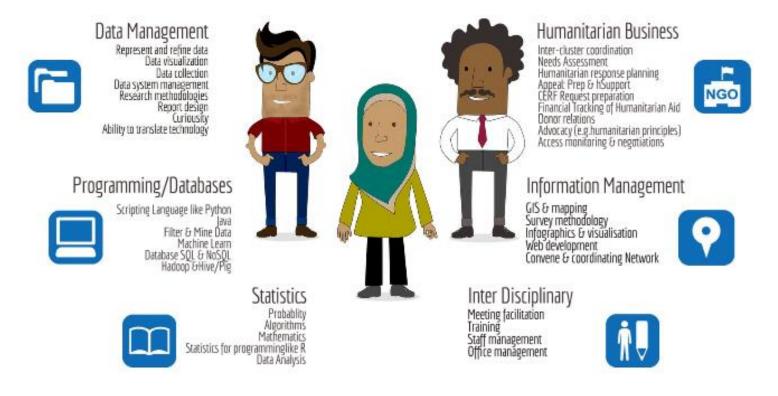
If foundations want the data to be able to compare the effectiveness of different kinds of programs, they need to fund that research specifically rather than hoping it will fall out of non-profits' work organically. And then the job of nonprofits should be only to show that they're well qualified to implement a particular kind of program, and not the validity of that kind of program itself.

Data for Social Good

- Amount and importance of data is increasing
- Most advanced use of data is for business and commercial purposes
- Social Organizations are trying to do good
- Hard to find the skills and resources to collect, understand and use data to make better decisions
- Data professionals with knowledge and skills want to become more engaged
- Worldwide 'Data for Social Good' movement

Data Science

Humanitarian Data Scientist Ever wondered about the skills of a data specialist during a crisis? And how to become one?



Online Courses: a collection of free courses to help you develop or sharpen your skills. http://goo.gl/JqDYtH





















Data for Good

Data for Good - Calgary

- Putting data into action for social good in Calgary
- Started in Nov 2013 and now have 600+ members
- Partnering to assist local non-profit organizations
 - DataThons: intense weekend events, significant preparation
 - DataCorps: project-based volunteer assistance to non-profits
- Arrange monthly Meetups and collaborate with organizations nationally & internationally

More Information

- Meetup Group: <u>www.meetup.com/Data-for-Good-Calgary/</u>
- Twitter: <u>@DataForGoodYYC</u>
- Geoff Zakaib: <u>geoff@lumina.ca</u>



Data for Good - Calgary

More

Home

Members Sponsors Photos Discussions



Calgary, AB Founded Nov 11, 2013

Data Do-Gooders	574
Group reviews	9
Past Meetups	20
Our calendar	Ē

9

Organizers:



Geoff Zakaib, Joy Robson, Victor

Anjos

Contact

If you have an interest in data and putting it into action for social good please join us at Data for Good - Calgary.

Join us!

* Who we are *

Data for Good inspired by DataKind.org is working for positive social action through "data in the service of humanity". If you are based in Calgary and have a desire to stay at the forefront of technological innovation & positive social change, please join us.

* What we do *

Data for Good brings together leading data scientists with high impact social organizations through a comprehensive, collaborative approach that leads to shared insights, greater understanding, and positive action through "data in the service of humanity". Data for Good leads a community of pioneering data scientists with the talent, commitment, and energy to open doors & inspire a new way of using the skills and tools of corporations & governments, to meet the needs of the NFP/NGO and social innovation sector.

* How we do it *

By organizing & hosting weekend DataThon events. A DataThon is a weekend event that matches up selected social organizations (that have well-defined data problems) with a team of volunteer data scientists to tackle their data-related challenges over a 24-48 hours period. The participants are fed throughout the weekend event & the results are presented to the social organizations at the end of the event. Some may refer to these types of events as "Hack-a-Thons", etc. These events are completely free for the participants and serve to energize, educate, and provide direct benefit to the NFP/NGO organizations, as well as to enlighten social sector groups about the power of being data-driven.

Other 'Data for Social Good' Organizations

- DataKind
- Data Analysts for Social Good, Do Good Data Conference
- Markets for Good
- Powered by Data
- Data Science for Social Good
- Data & Society Research Institute

Datathon - Calgary Distress Centre

- Internal data from Crisis Line and 211
- External data on weather, suicide statistics, etc.
- Objectives were formulated and grouped
 - Analytics and visualization
 - Spatial analysis
 - Text analytics

• Very successful DataThon held in May 2015

- Insights generated and documented
- Results used in discussions with key stakeholders
- New data analyst role justified and placed in 2016 budget



Distress Centre – Distress is Complex

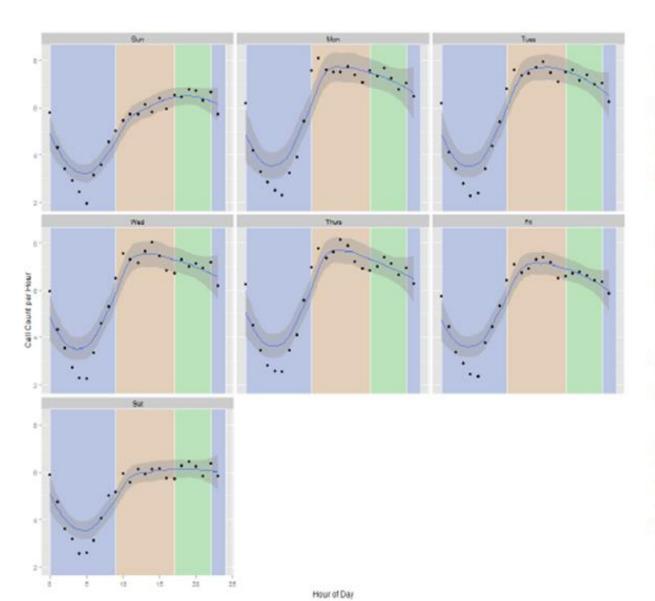
179,975 calls Dec 5, 2011 - Jan 11, 2015	Abuse Violence	Addictions	Family	Financial Economical	Mental Health	Legal Miscellaneous	Physical Health	Relationships	Sexual
Abuse Violence	14	2	6	3	12	1	4	7	0
Addictions	2	9	3	2	7	1	3	4	0
Family	6	3	26	7	22	2	8	14	0
Financial Economical	3	2	7	20	16	2	5	9	0
Mental Health	12	7	22	16	72	4	18	30	1
Legal Miscellaneous	1	1	2	2	4	17	1	2	0
Physical Health	4	з	8	5	18	1	22	9	0
Relationships	7	4	14	9	30	2	9	36	1
Sexual	0	0	0	0	1	0	0	1	1

Nine major categories of issues, each with a large variety of detailed categories

- 72% Mental Health
- 36% Relationships
- 26% Family
- 22% Physical Health
- 20% Financial
- 17% Legal/Misc
- 14% Abuse, Violence
 - 9% Addictions
- 1% Sexual

All calls have more than one issue

Distress Centre – Call Volume is Cyclical

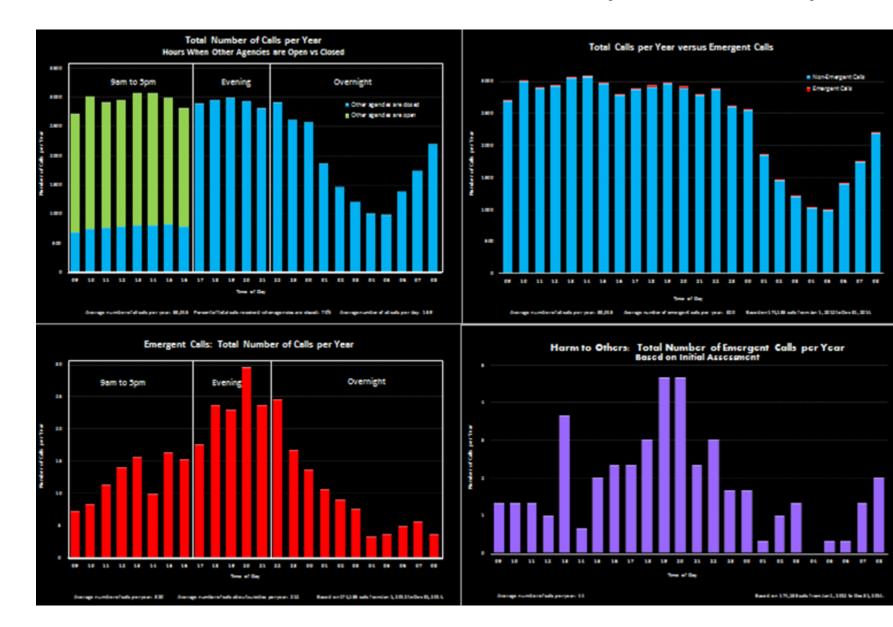


Daily volume pattern changes during the hour of the day and day of the week

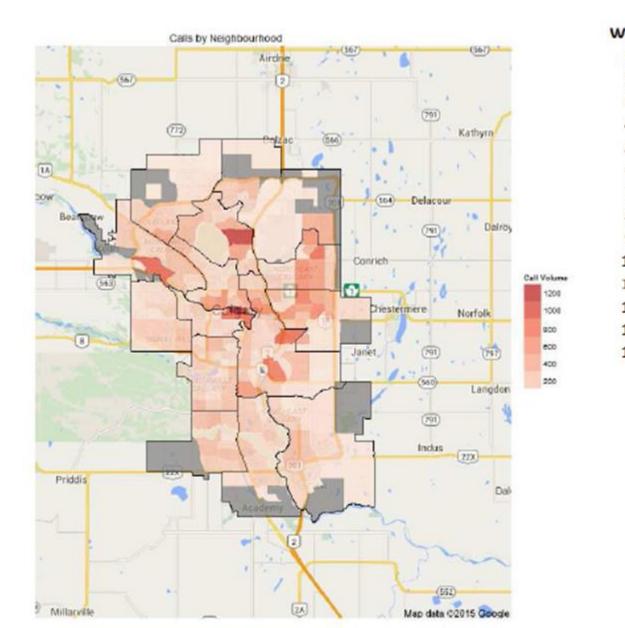
Call volume picks up after 9:00, peaks at 1:00-2:00 and gradually decreases until 11:00 pm

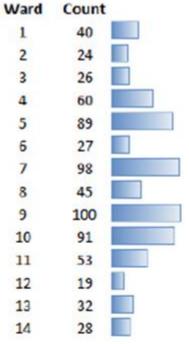
Weekends volume is lower and peaks less during the day

Distress Centre – Time-of-Day Call Analysis

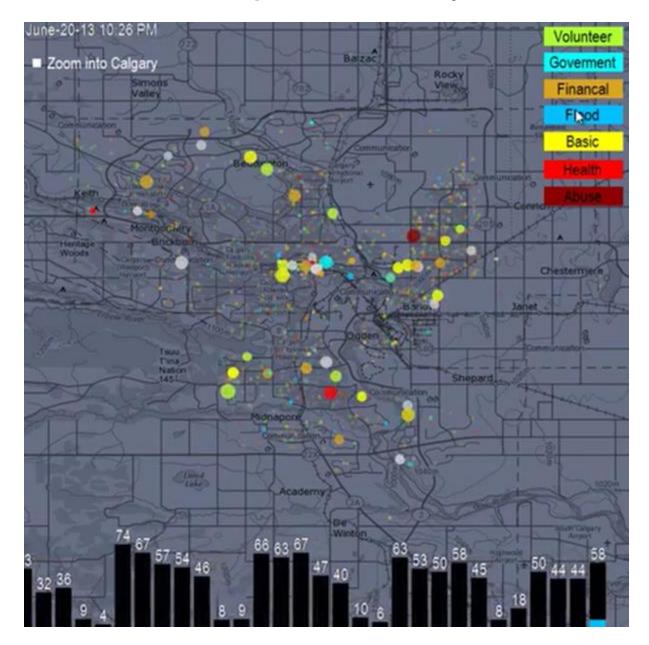


Distress Centre – Call Volume across City





Distress Centre – Spatial Analysis of 211 Calls



Data for Good and Partner Organizations









CAWST









Canadian Mental Health Association Edmonton