Data Management at the University of Waterloo

Colin Bell
Director, Enterprise Architecture
University of Waterloo

cpbell@uwaterloo.ca

May 10, 2016

Presented at the 2016 Saskatchewan Connections Conference

Agenda

- Legislation
- IPC Design Principles
- Data Management Principles
- Data Management at Waterloo

Data Management in Ontario Higher Education

- The Office of the Information and Privacy Commissioner (IPC) is responsible for ensuring compliance with Ontario's access and privacy laws. https://www.ipc.on.ca

- Freedom of Information and Protection of Privacy Act (FIPPA), R.S.O. 1990, c. F.31

http://bit.ly/1FmppGD

Privacy 101

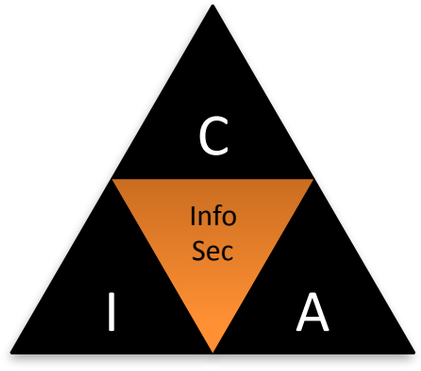
- Information Privacy is the right of an individual to exercise control over the collection, use, disclosure and retention of his or her personal information, including his or her student records.
- It is a legal matter.
- It is not an IT matter.
- Privacy ≠ Secrecy

Privacy Principles

1	**Accountability**
2	**Identifying Purposes**
3	**Consent**
4	**Limiting Collection**
5	**Limiting Use, Disclosure and Retention**
6	**Accuracy**
7	**Safeguards**
8	**Openness**
9	**Individual Access**
10	**Challenging Compliance**

No Privacy without Security

- Information Security - CIA Triad:



- **Confidentiality** - restricted access is enforced and maintained as expected.
- **Integrity** - assurance information can be trusted.
- **Availability** - access is possible when required.

Access 101

- Underlying Concept:
  - Citizens need information to take part in democratic process.
  - Information should be shared widely to those who need access (while respecting privacy and confidentiality).
  - Individual Privacy is a core value to the public.

Access 101 (cont'd)

- Through Freedom of Information (FOI) Requests, the public has a right to access records in an institution's custody or control unless:
  - an exemption applies; or
  - it is determined that the request is frivolous or vexatious; or
  - the information is excluded from Legislation.

Access 101 (cont'd)

- Must sever and release non-exempt portions and provide reasons for exemption.

- Directory of Records required to list all available records and personal information banks.

- All decisions made can be reviewed by the Information & Privacy Commissioner (IPC).

Routine Disclosure

"In the spirit of the Act, unless there is a statutory requirement or reason not to release the information, routine disclosure should become the norm."

Information/Privacy Commissioner Guidelines on Routine Disclosure

Hot Side Hot
Cool Side Crisp

McDLT

Hot Side Hot
Cool Side Crisp

Confidential

Public

Access by Design

www.privacybydesign.ca

McDLT

IPC: Privacy by Design (PbD)

1. **Proactive** not Reactive; **Preventative** not Remedial
2. Privacy as the **Default Setting**
3. Privacy **Embedded** into Design
4. Full Functionality — **Positive-Sum**, not Zero-Sum
5. End-to-End Security — **Full Lifecycle Protection**
6. **Visibility** and **Transparency** — Keep it **Open**
7. **Respect** for User Privacy — Keep it **User-Centric**

http://bit.ly/1OfkUTo

IPC: Access by Design (AbD)

1. **Proactive**, not Reactive
2. Access **Embedded** into Design
3. Openness and Transparency = **Accountability**
4. Fosters **Collaboration**
5. Enhances **Efficient Government**
6. Makes Access Truly **Accessible**
7. Increases **Quality** of Information

http://bit.ly/1fYrWQq

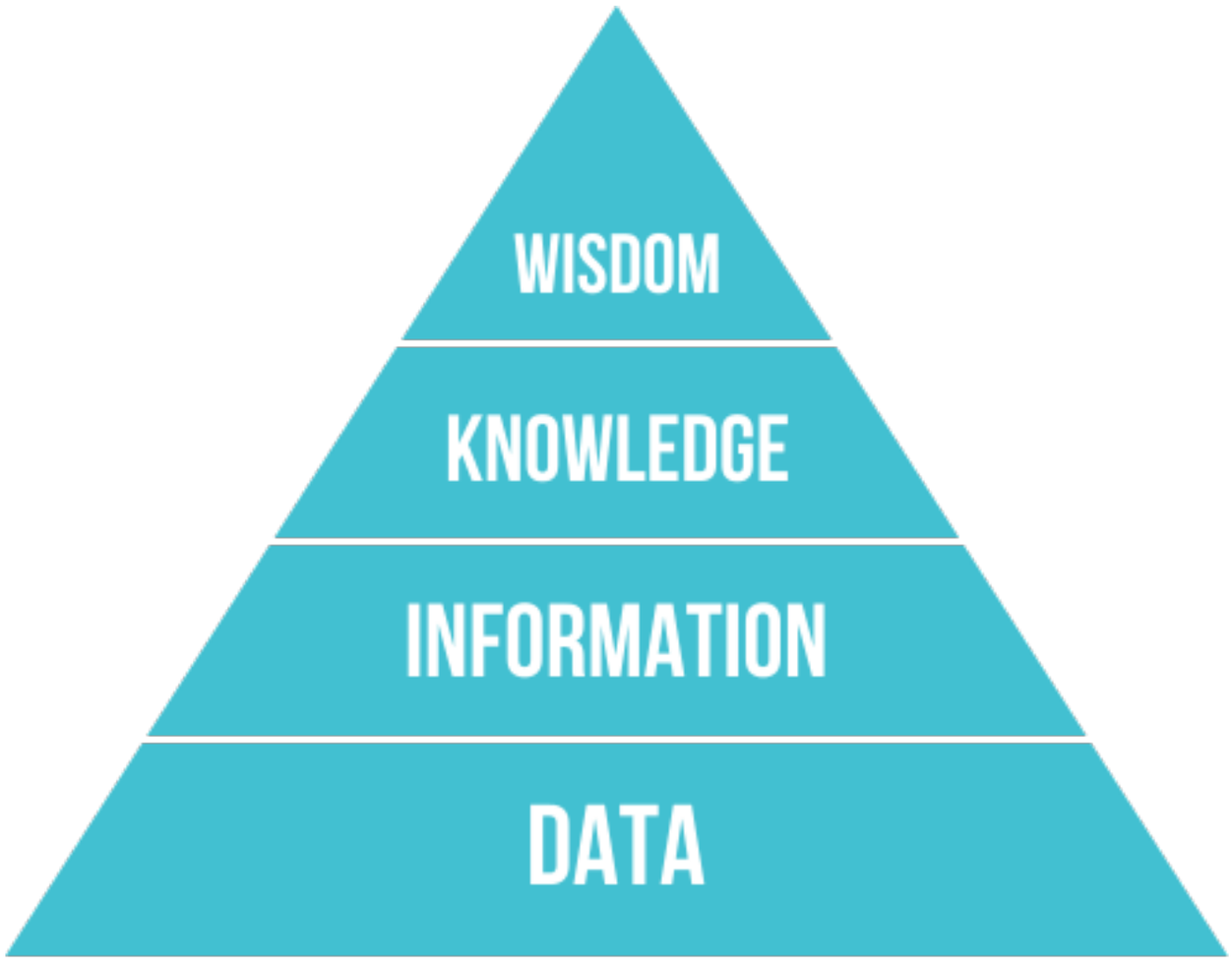
Summary

- We are in control and custody of information.
- Some of that information is confidential.
- Some of that information is public.
- The IPC recommends public information should be subject to routine disclosure.
- The IPC provides design principles (PbD, AbD) to guide us as we design business processes and technology.

DIKW

Data, Information, Knowledge, Wisdom

- **Data** – Facts without context.
- **Information** – Data “in formation”, data in context.
- **Knowledge** – Information in context. The ability to answer questions based on information; to know how something happens.
- **Wisdom** – Being able to act (based on knowledge) to amplify positive/dampen negative outcomes.



"DIKW Pyramid" by Longlivetheux - Own work. Licensed under CC BY-SA 4.0 via Commons
https://commons.wikimedia.org/wiki/File:DIKW_Pyramid.svg#/media/File:DIKW_Pyramid.svg

http://bit.ly/1M9qaoC

DIKW

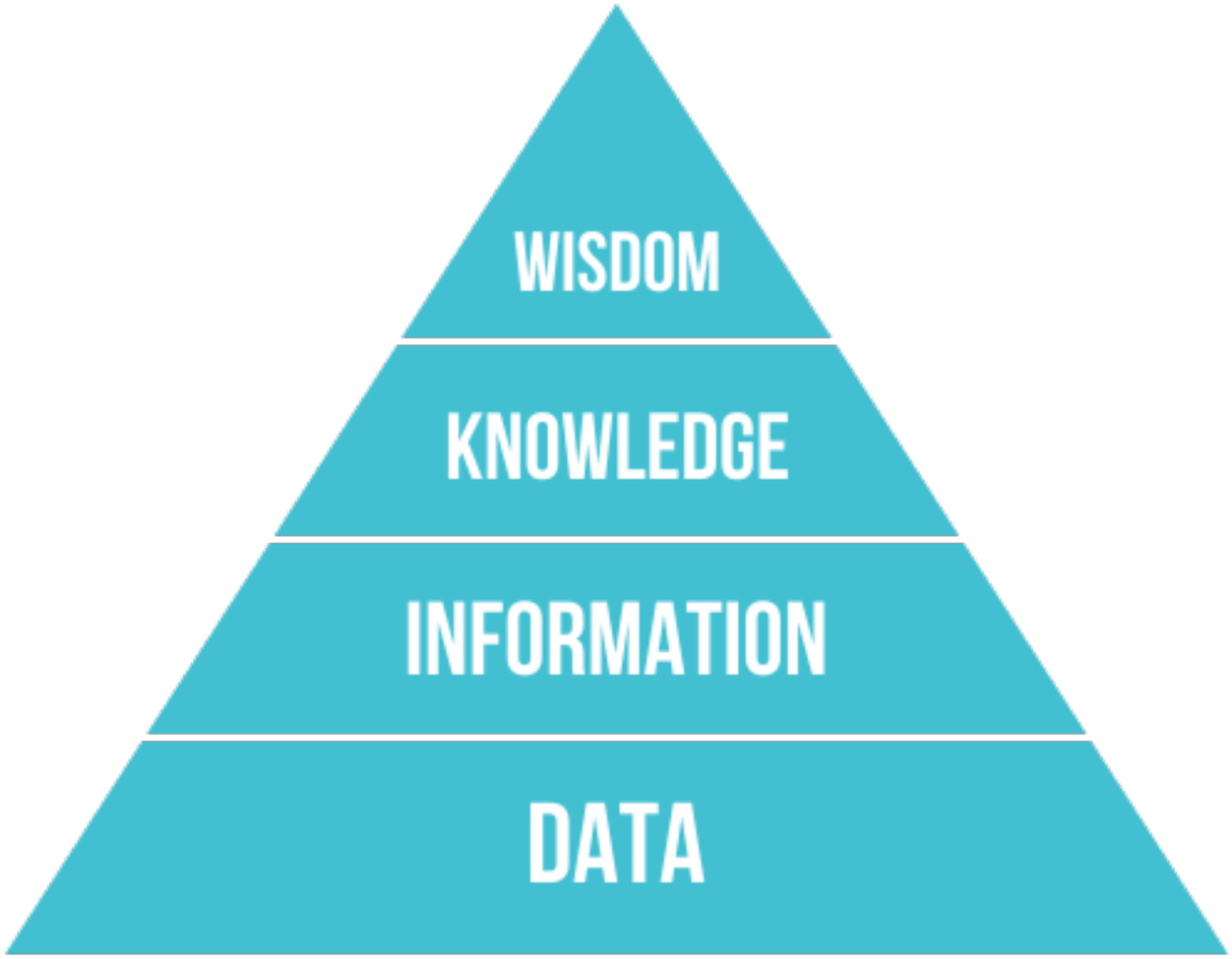
Data, Information, Knowledge, Wisdom

- **Data** – 24

- **Information** – Temperature in office: 24°C

- **Knowledge** – The air conditioner should prevent the temperature from rising over 21°C

- **Wisdom** – The air conditioner requires maintenance.

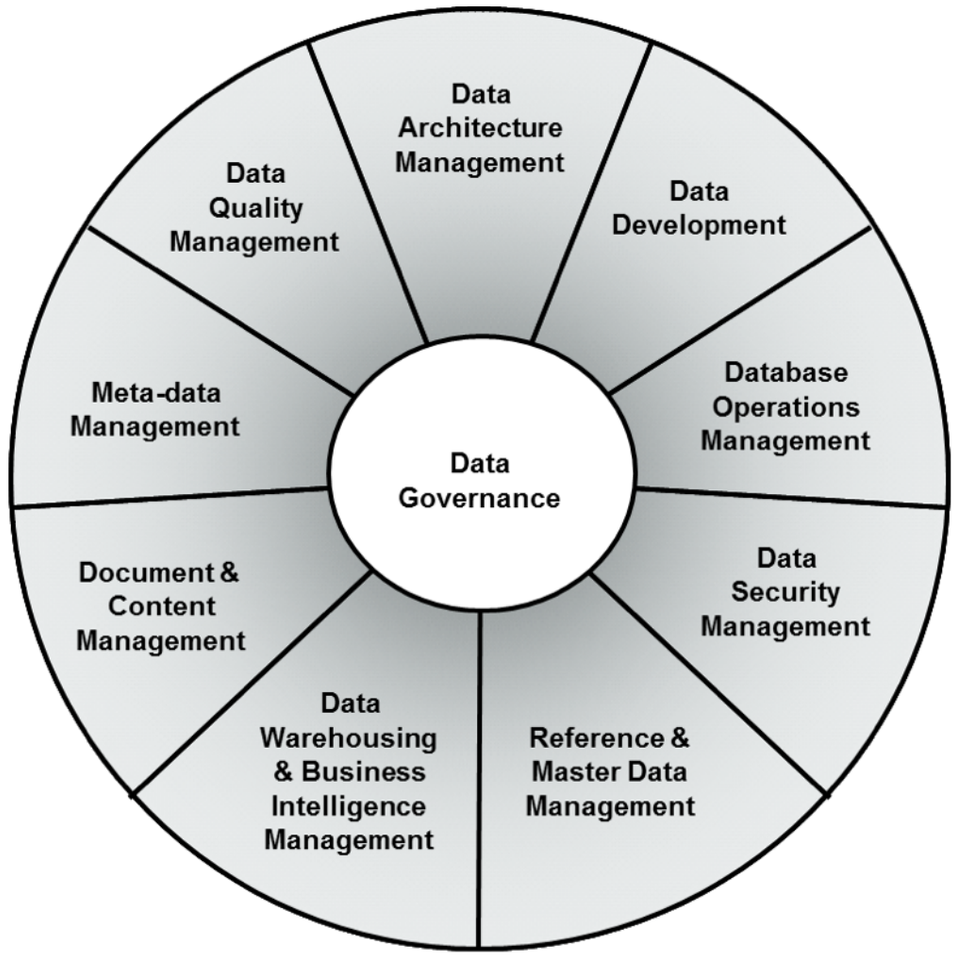


"DIKW Pyramid" by Longlivetheux - Own work. Licensed under CC BY-SA 4.0 via Commons
https://commons.wikimedia.org/wiki/File:DIKW_Pyramid.svg#/media/File:DIKW_Pyramid.svg

http://bit.ly/1M9qaoC

Data Management

- DAMA DMBOK
  - Data Governance
  - Data Architecture Management
  - Data Development
  - Data Operations Management
  - Data Security Management
  - Reference and Master Data Management
  - Data Warehousing and Business Intelligence
  - Document and Content Management
  - Meta-data Management
  - Data Quality Management



http://bit.ly/1KuHRV4

Data Management

- **Data Governance** – planning, supervision and control over data management and use

- **Data Architecture Management** – connection of data to the larger Enterprise Architecture strategy

- **Data Development** – analysis, design, building, testing, deployment and maintenance

- **Database Operations Management** – support for structured physical data assets

- **Data Security Management** – ensuring privacy, confidentiality and appropriate access

http://bit.ly/1KuHRV4

Data Management

- **Reference & Master Data Management** – managing golden versions and replicas

- **Data Warehousing & Business Intelligence Management** – enabling access to decision support data for reporting and analysis

- **Document & Content Management** – storing, protecting, indexing and enabling access to data found in unstructured sources (electronic files and physical records)

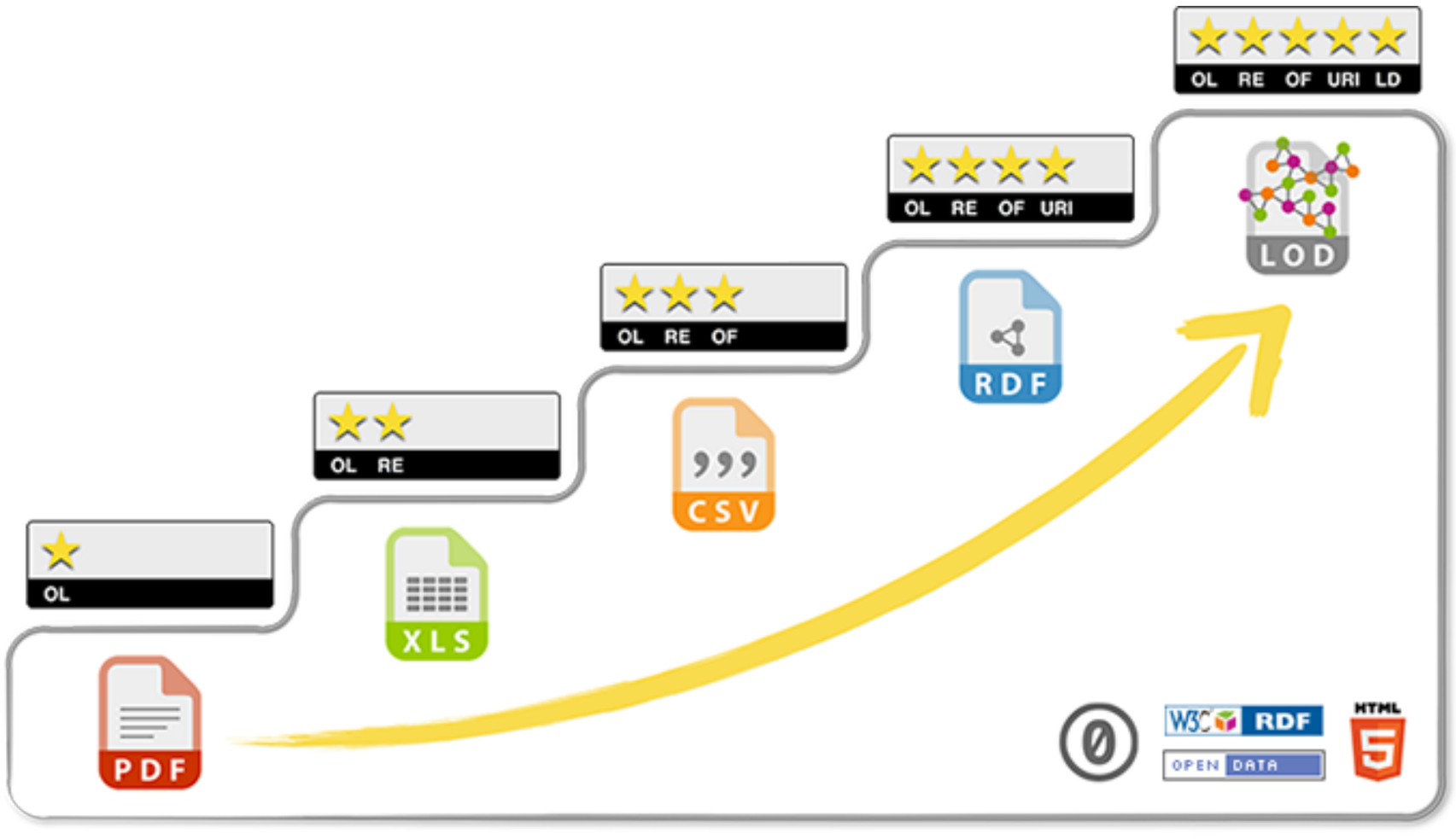
- **Meta Data Management** – integrating, controlling and delivering meta data

- **Data Quality Management** – defining, monitoring and improving data quality

http://bit.ly/1KuHRV4

5-Star Data

- Tim Berners-Lee Open Data Model



http://5stardata.info/en/

5-Star Data

★	Available on the web (whatever format) *but with an open licence, to be Open Data*
★★	Available as machine-readable structured data (e.g. excel instead of image scan of a table)
★★★	as (2) plus non-proprietary format (e.g. CSV instead of excel)
★★★★	All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
★★★★★	All the above, plus: Link your data to other people's data to provide context

http://bit.ly/21MR3Zt

Example: Student Portal

https://uwaterloo.ca/student-portal/

"I would like to have mapping and navigation features available for on campus."

GO

Student Success Office

Student Portal

Secure Information Exchange (PbD / PII)

Open Information Exchange (AbD / Open Data)

Learn Data

Student Accounts Data

Student Financial Aid Data

Watcard Data

FEDS Data

Plant Ops Data

RO Data

GSO Data

Food Services Data

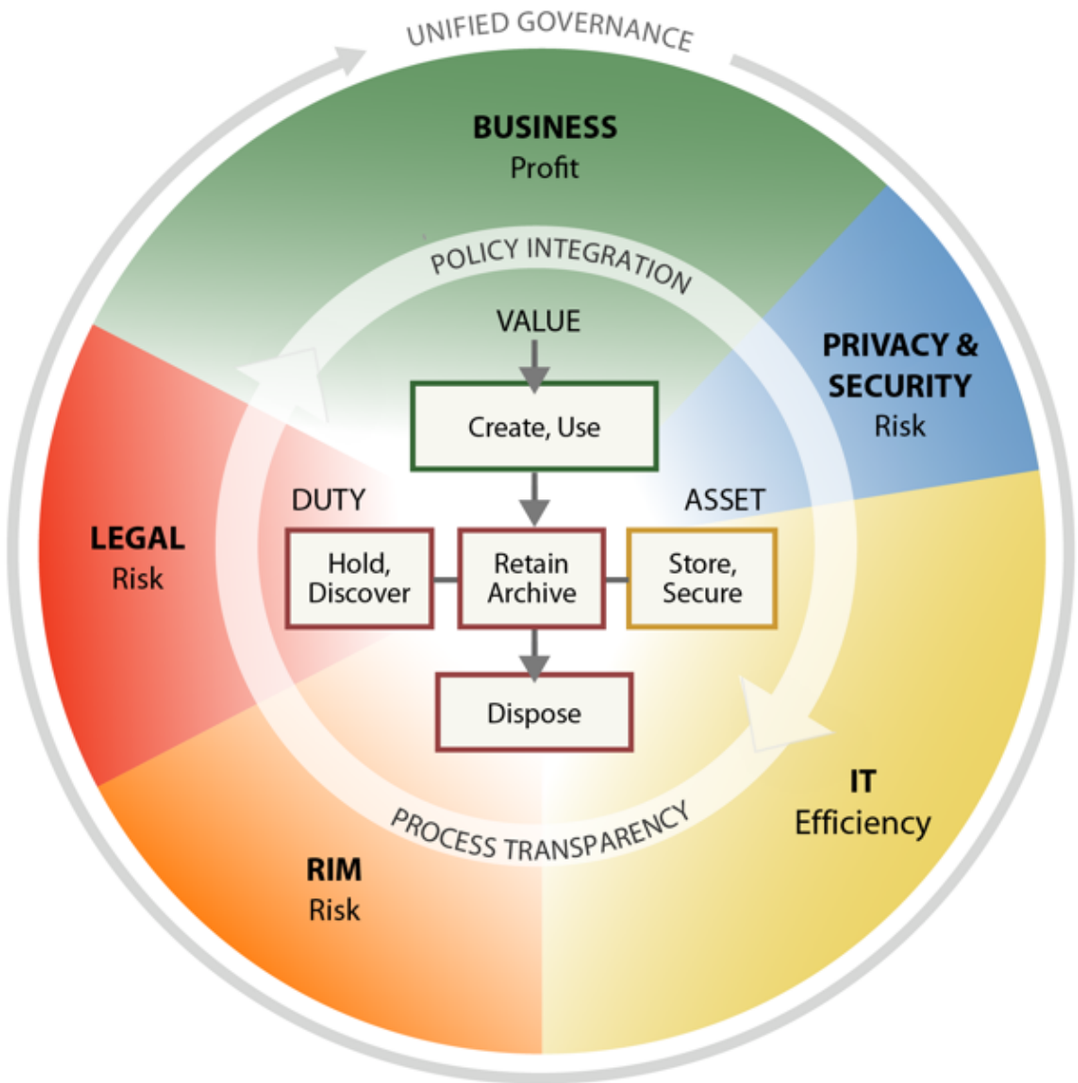
CECA Data

CPA (WCMS) Data

Governance

Information Governance Reference Model (IGRM)

Linking duty + value to information asset = efficient, effective management



Duty: Legal obligation for specific information

Value: Utility or business purpose of specific information

Asset: Specific container of information

http://bit.ly/1jDqgia

Data Management at Waterloo

- Core “Business” Data
  - Finance
  - Human Resources
  - Student Information
  - Research Information

Data Management at Waterloo

- Core “Business” Data
  - Finance
  - Human Resources
  - Student Information
  - Research Information
- Other Data
  - Research Data
    - On People
    - Corporate Partner Owned
  - Healthcare Data

Data Management at Waterloo

- Core “Business” Data
  - Finance
  - Human Resources
  - Student Information
  - Research Information
- ~~Other Data~~
  - ~~Research Data~~
    - ~~On People~~
    - ~~Corporate Partner Owned~~
  - ~~Healthcare Data~~

Data Management at Waterloo

Business Data

OLTP — Online Transaction Processing

ODS — Operational Data Stores

OLAP — Online Analytical Processing

Confidential

ANON OLAP — Anonymized* OLAP

Public

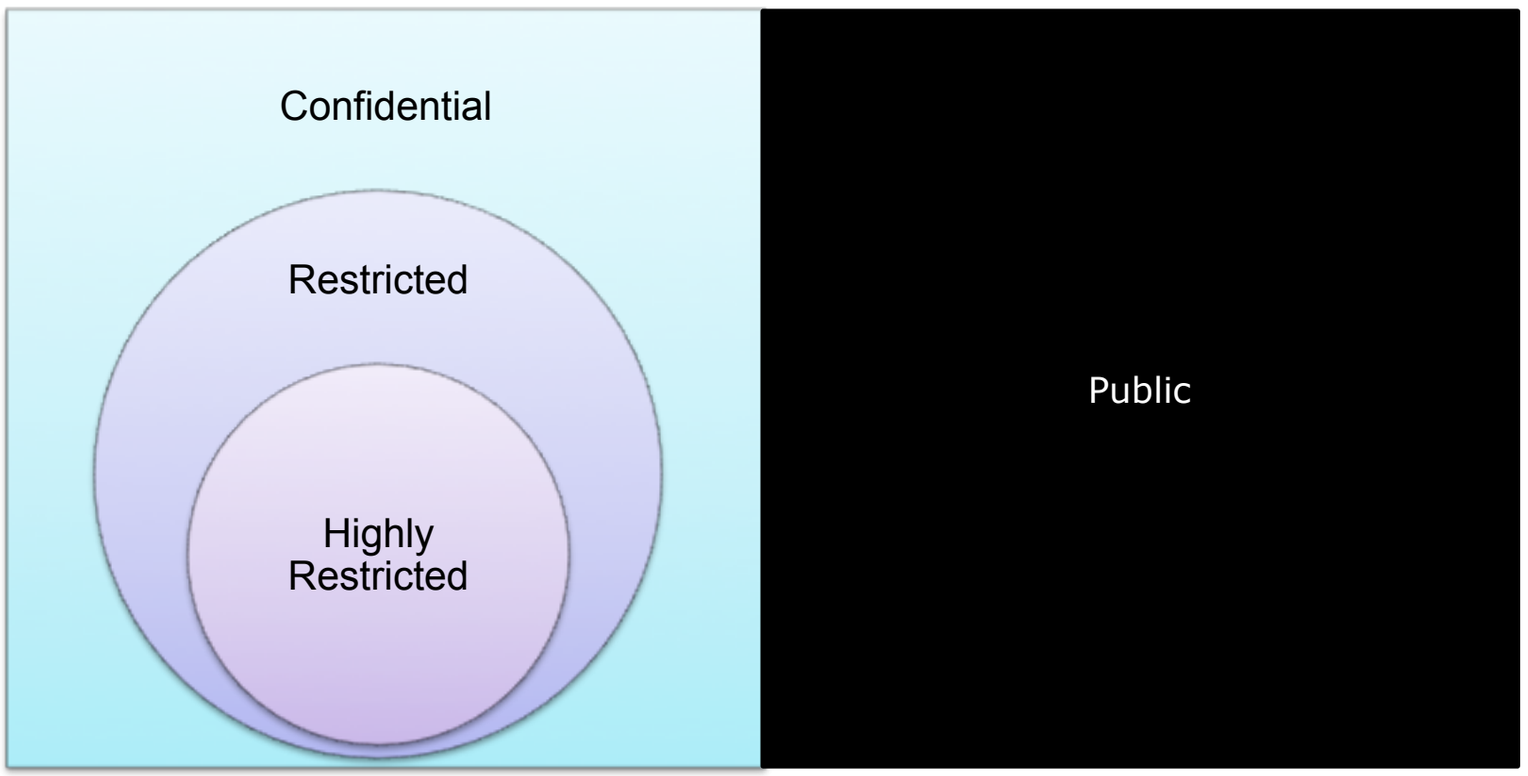
* http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf

Data Management at Waterloo

- Strengths
- Weaknesses
- Opportunities
- Threats

Waterloo Strengths

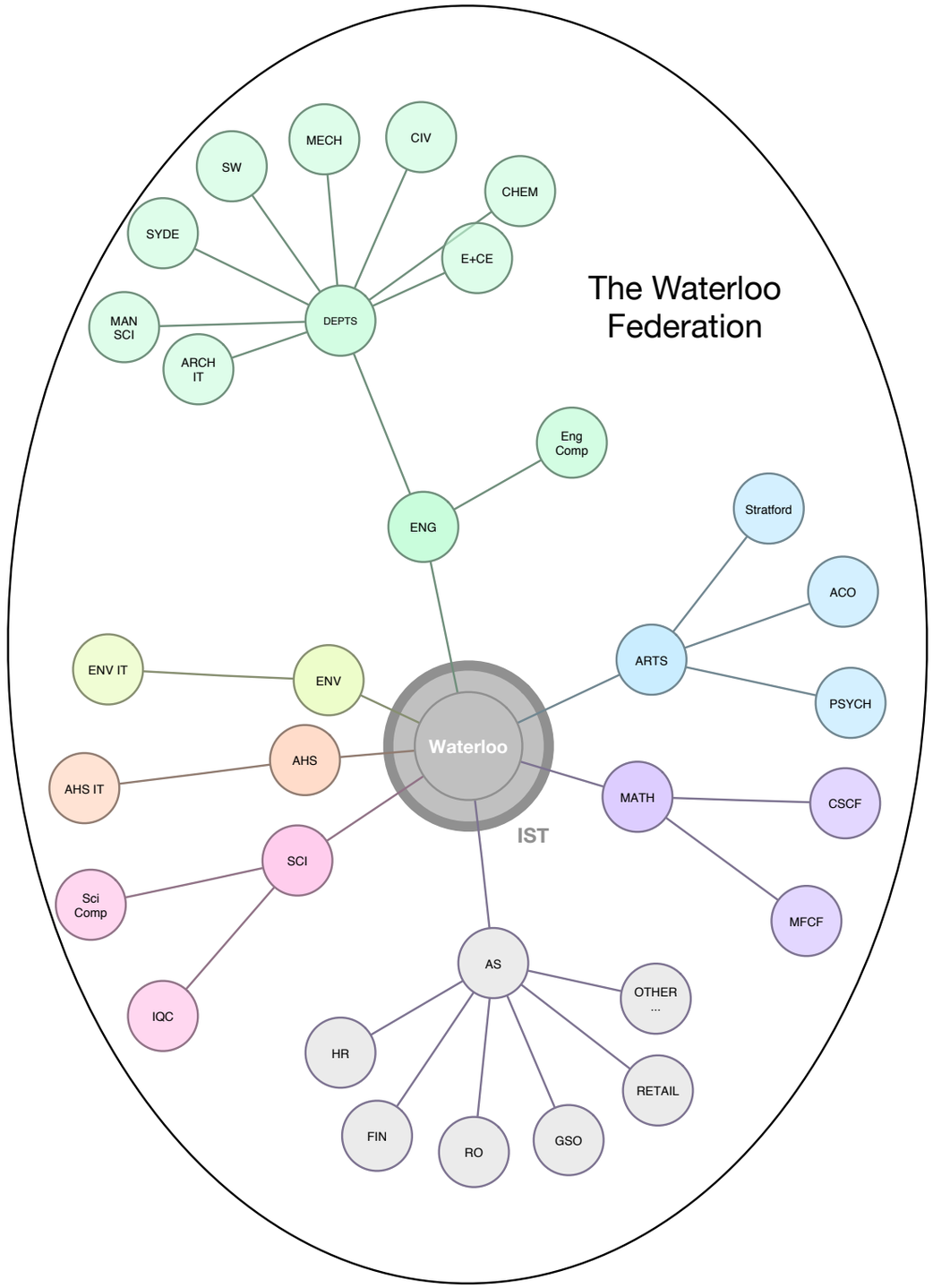
- Policy 8 - Information Security
  - Roles + Responsibilities (Steward >> Custodian >> User)
  - Information Security Classifications



https://uwaterloo.ca/secretariat-general-counsel/policies-procedures-guidelines/policy-8

Waterloo Strengths

- Decentralized
- Federated Organization



Waterloo Strengths

- Statement on Information Management
- Information is “Vital Asset”
- Recognize information’s part in governance, administration, service planning and delivery, and performance management of University.

Waterloo Strengths

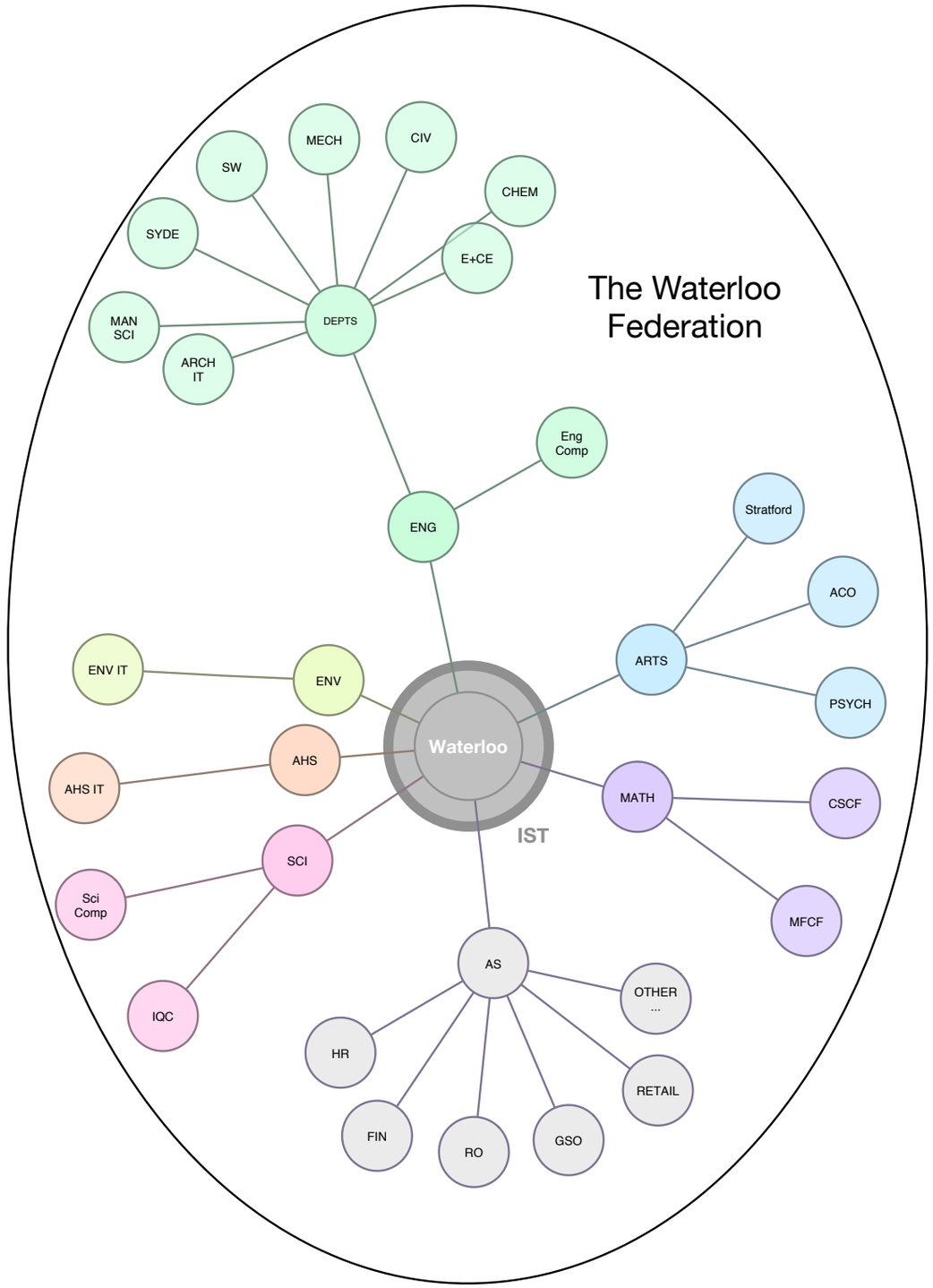
- Open Data is well established and growing.
    - https://api.uwaterloo.ca
- Shift to Confidential Data is underway.
- Moving up the 5-Star Data Model.

Waterloo Weaknesses

- Policy 8 - Information Security
- AND …
    - Policy 11 - University Risk Management
    - Policy 12 - Records Management
    - Policy 13 - Archives
    - Policy 19 - Access to and Release of Student Information
    - Others?

Waterloo Weaknesses

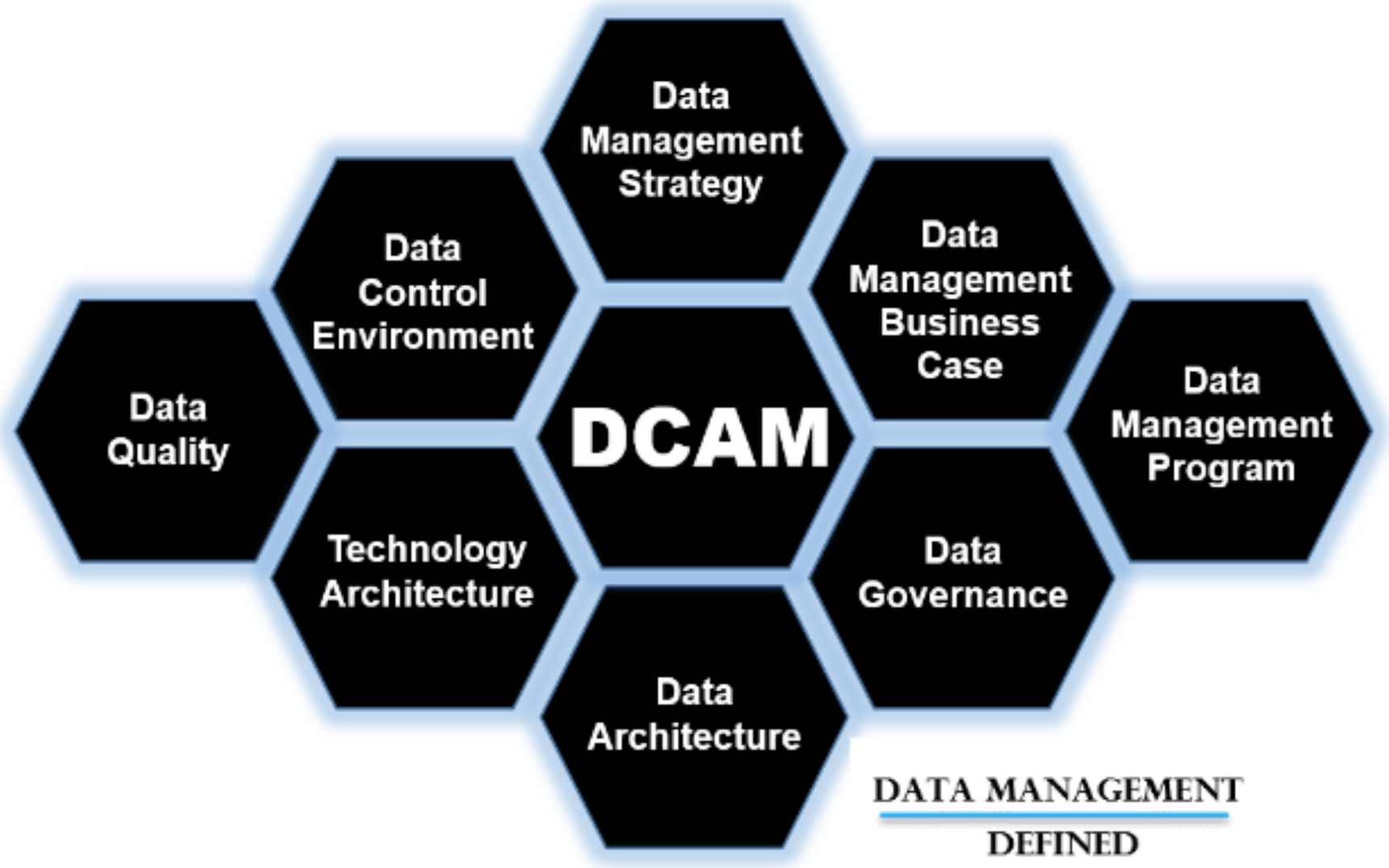
- Decentralized
- Federated Organization



Waterloo Opportunities

- EDM Council (http://edmcouncil.org) Membership
  - Financial firms really care about their data.
  - Data Management Capability Assessment Model (DCAM) provides tool to measure state.
  - Not a perfect fit for Higher Ed, but we are using their tools as a starting point.

Waterloo Opportunities



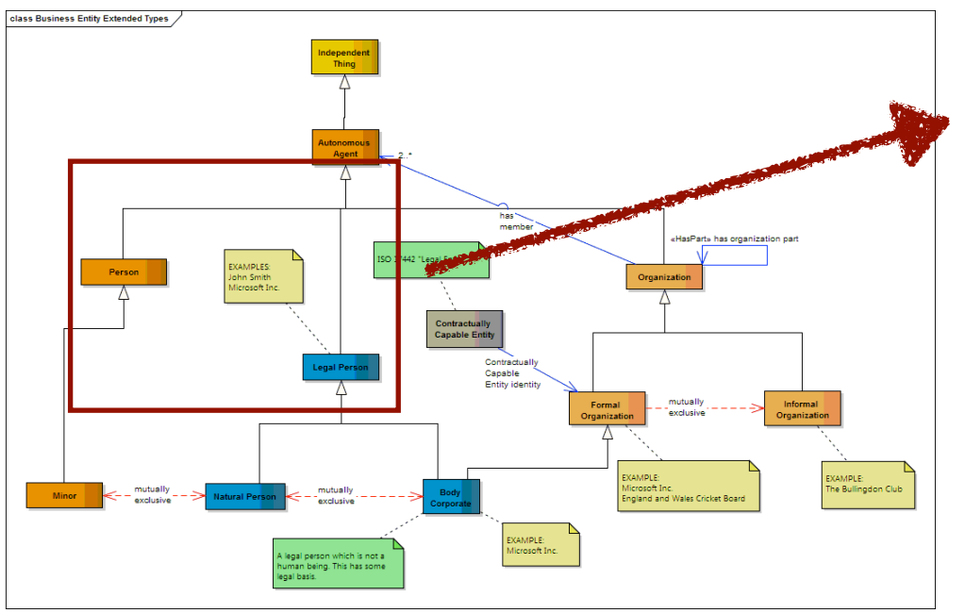
DATA MANAGEMENT DEFINED

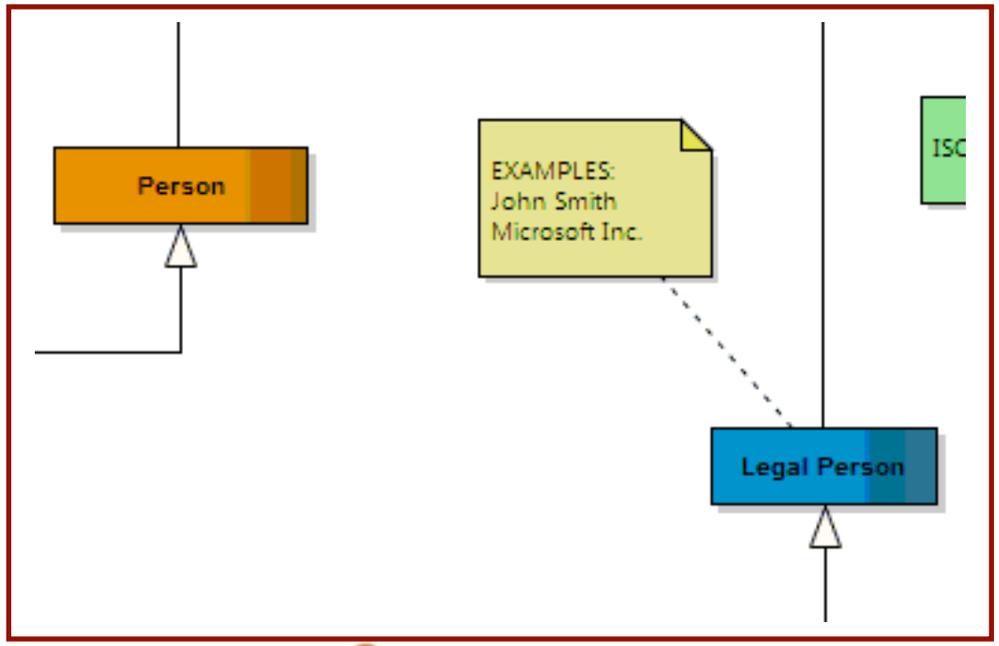
ENTERPRISE DATA MANAGEMENT
EDM COUNCIL

DCAM
Data Management Capability Assessment Model

Waterloo Opportunities

- Learning from EDM Council (http://edmcouncil.org)
  - “Big Data” data quality and integration problems are being solved with Web 3.0 “Semantic Web” technologies.



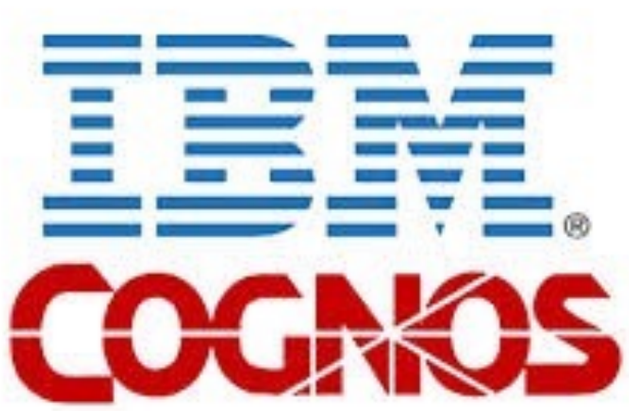


ENTERPRISE DATA MANAGEMENT
EDM
COUNCIL

FIBO
Financial Industry Business Ontology

Waterloo Opportunities

- Innovation in the decentralized areas brings multiple tools onto campus.



Waterloo Threats

- Multiple tools in distributed use lead to non-standard results.
    - NEED: Accepted and shared data warehousing.
    - NEED: Documented metadata + taxonomies.
- More often than not data is ‘copied’ into local spreadsheets. “It’s easier.”
- We need rock solid data to make decisions.

Questions + Answers

- Or comments… how do your organizations deal with data management?

Thank you!

Injury Time:
Enterprise Architecture