



Challenges of Preserving Digital Twins

I Trust AI International Symposium
Teatro Municipal de Tinajo
Canary Islands
Oct 27, 2022

Tracey P. Lauriault (remote)
Associate Professor, Critical Media & Big Data
School of Journalism & Communication
Carleton University, Ottawa (ON) Canada
orcid.org/0000-0003-1847-2738
Tracey.Lauriault@Carleton.ca
[@TraceyLauriault](https://twitter.com/TraceyLauriault)

I Trust AI



I Trust AI aims to:

- Generate new knowledge on the uses of AI, such as Machine Learning (ML), on public records
- Improve upon existing tools and create new ML tools that will address:
 - archival needs, such as machine translation, image recognition and description, optical character recognition (OCR) and handwritten text recognition, text summarization and classification, and text style transfer for language civilization (e.g., removal of bias, hate, and sexism).
 - The tools developed will be tailored to, tested for, and deployable by the records and archives community
- **Produce best practices, standards, and guidelines for applying ML tools across an entire problem space**, bringing archival knowledge and practice to bear on problems such as bias in ML, explainable artificial intelligence (XAI), and image description
- **Enrich research and lead to knowledge co-creation in several disciplines**, including archival science, records management, AI, cybersecurity, information science, law, and ethics, through knowledge exchange and uptake between scholars and practitioners within and among those discipline
- Train a substantial cohort of students -- future scholars and professionals -- who will bring their enhanced knowledge to the institutions, organizations, communities, and governments they will serve, and to the archives and records community as a whole



I Trust AI Research Objectives:

1. Identify specific AI technologies that can address critical records and archives challenges
2. Determine the risks and benefits of using AI technologies on records and archives
3. Establish how archival concepts and principles can inform the development of responsible AI
4. Validate outcomes from Objective 3 through case studies and demonstrations



Framing & Methodology

Critical Data Studies

- Data do not exist independently of the ideas, instruments, practices, contexts and knowledges used to generate, process and analyze them
- Application of social theory to the study of data and technology
- Conceptualizing Data & Technological Systems



Today's Data are Part of Tomorrow's Research: Archival Issues in the Sciences*

TRACEY P. LAURIAULT, BARBARA L. CRAIG, D.R. FRASER TAYLOR, and PETER L. PULSIFER

RÉSUMÉ Les données scientifiques sont essentielles à la formation en sciences et à la prise de décision éclairée au sujet de la santé, de l'environnement et de l'économie. Les ensembles de données cumulatives aident à comprendre les tendances, les fréquences et les courants, et ils peuvent servir de base pour développer des prévisions. Cet article se penche sur la préservation des données scientifiques et des portails

1 Toward Critical Data Studies

Charting and Unpacking Data Assemblages and Their Work

Rob Kitchin and Tracey P. Lauriault

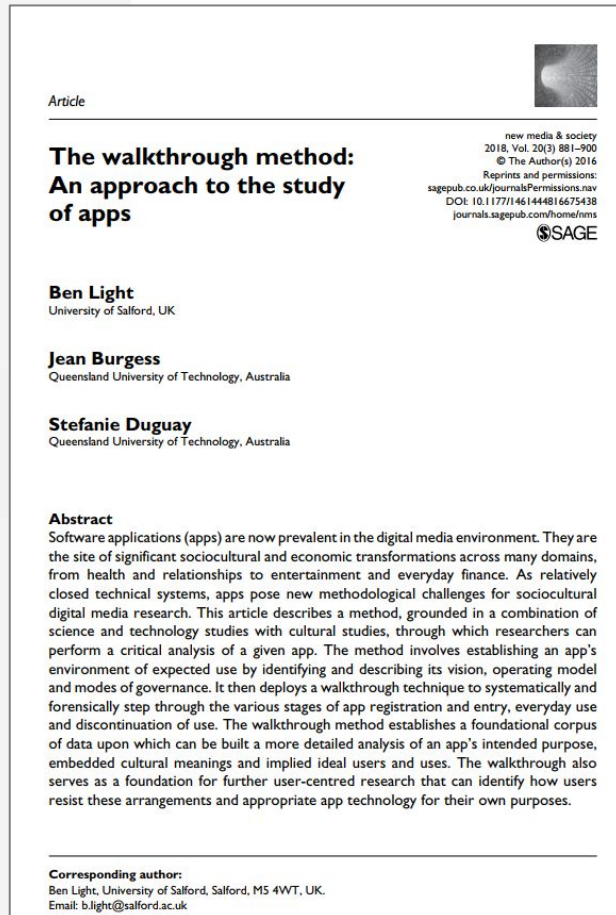
A Critical Approach to Data

Societies have collected, stored, and analyzed data for a couple of millennia as a means to record and manage their activities. For example, the ancient Egyptians collected administrative records of land deeds, field sizes, and livestock for taxation purposes, the Domesday Book in 1086 captured demographic data, double-entry bookkeeping was used by bankers and insurers in the fourteenth century, and the first national registry was undertaken in Sweden in the seventeenth century.¹ It was not until the seventeenth century, however, that the term "data" was used for the first time in the English language, thanks to the growth of science, the development of statistics, and the shift from knowledge built from theology, exhortation, and sentiment to facts, evidence, and the testing of theory through experiment.² Over time the importance of data has grown, becoming central to how knowledge is produced, business conducted, and governance enacted. Data provide the key inputs to systems that individuals, institutions, businesses, and the sciences employ in order to understand, explain, manage, regulate, and predict the world we live in and are used to create new innovations, products, and policies.

The volume, variety, and use of data have grown enormously since the seventeenth century, and there has long been the creation and maintenance of very large data sets, such as censuses or government administrative and natural resource databases. Such databases, however, have typically been generated every few years or are sampled. In contrast, over the

domaines, en ciblant la qualité des données authentité – et en examinant comment ces données. Les auteurs donnent des définitions perspectives à la fois scientifiques et archivistes de la littérature sur le sujet (publications nationales et internationales, d'organismes financement, ainsi que des observations cas d'InterPARES 2 et de « General Study années scientifiques), cet article examine naissances » électronique (« machine-base » répercussions possibles sur la préservation et particulier sur les ontologies formelles. Il us le contexte d'un environnement Web 2.0, ntifiques, et le fait que ce domaine ne figure ent. Les auteurs avancent que les archivistes iques créateurs de données afin de compren e données sont des mécanismes dont les e leurs pratiques de préservation; et que ce

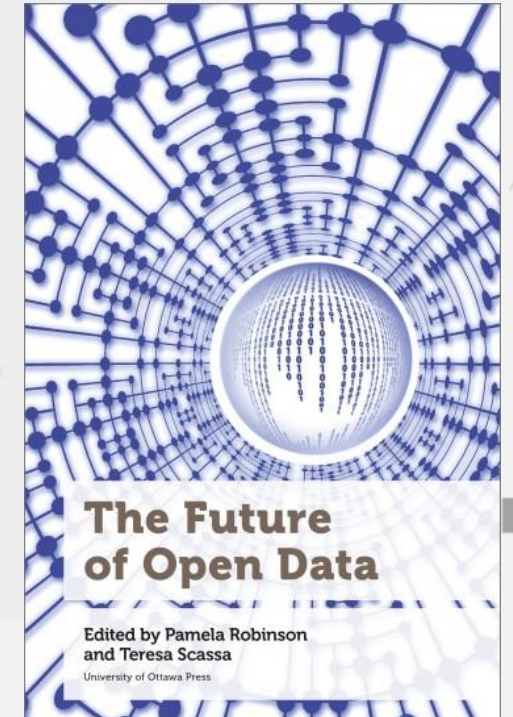
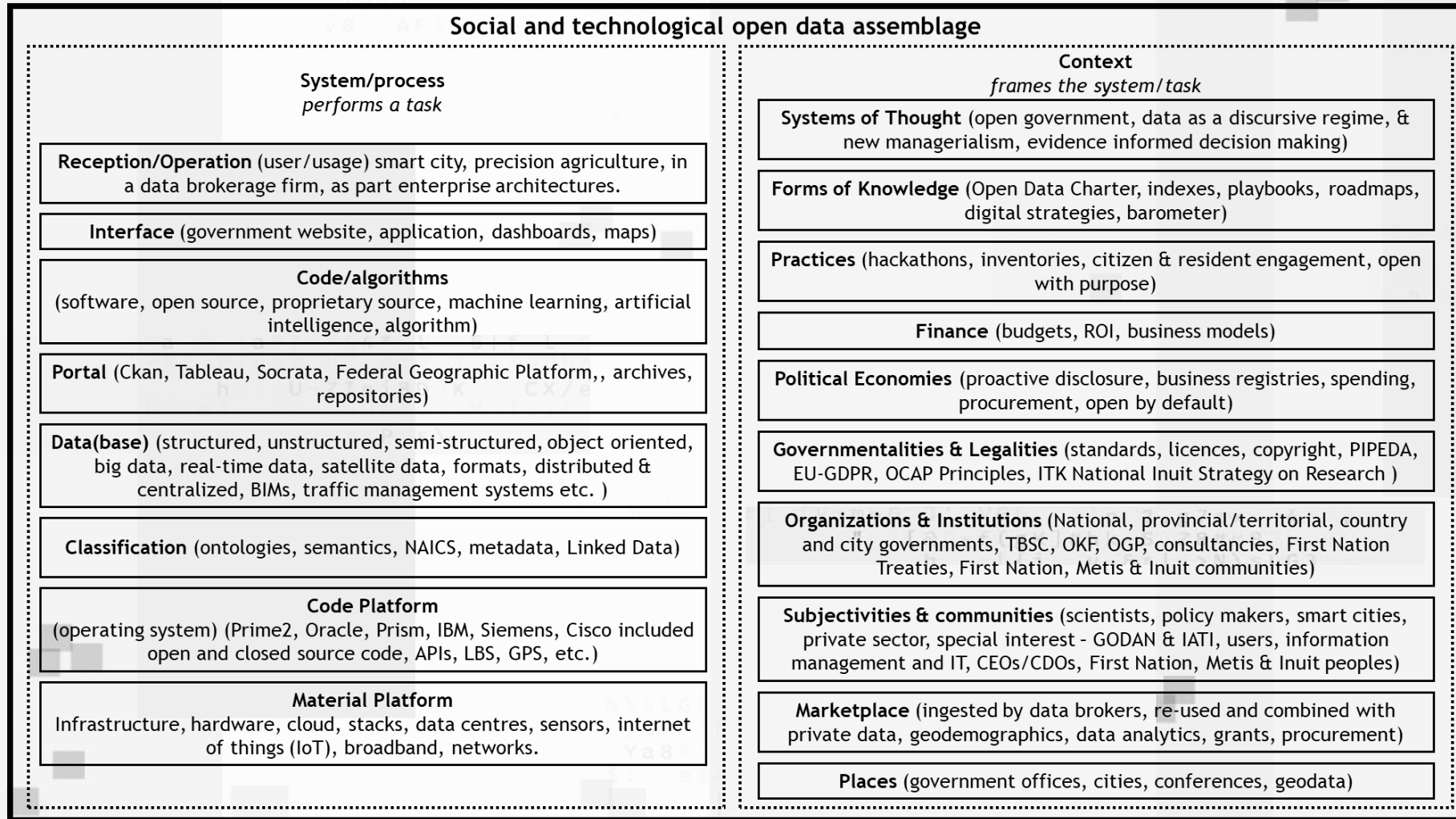
Walkthrough Method



Engage directly w/ a technological system to examine its technological mechanism & embedded cultural references

1. Environment of expected use
 - Vision, operating model, modes of governance
2. Walkthrough technique
 - Systematically & forensically step through the stages of use
 - Build a corpus of knowledge
3. Analysis
 - intended purpose, embedded cultural meaning, implied ideal use & users
 - How users resist intended purposes & use the technological system differently

Social & Technological Data Assemblage



Lauriault, T. P.
Looking Back Toward A
“Smarter” Open Data
Future

Methodology



InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records

APPENDIX 5

23 Case Study Questions that the researchers should be able to answer at the completion of their investigation

APPENDIX 8

InterPARES 2 Reporting Framework

23 Case Study Questions that the researchers should be able to answer at the completion of their investigation

March 18, 2003

1. What activities of the creator have you investigated?
2. Which of these activities generate the digital entities that are the focus of your investigation?
3. For what purpose(s) are the digital entities you have examined?
4. What form do these digital entities take? (e.g. e-mail, CAD)
 - 4a. What are the key formal elements, attributes, and behaviors of these entities?
 - 4b. What are the digital components of which they consist?
 - 4c. What is the relationship between the intellectual aspect and the technical aspect of these entities?
 - 4d. How are the digital entities identified (e.g., is there a unique identifier)?
 - 4e. In the organization of the digital entities, what kind of structure is used?
 - 4f. What determines the way in which the digital entities are organized?
5. How are those digital entities created?
 - 5a. What is the nature of the system(s) with which they are created? (e.g. software, hardware, peripherals etc.)
 - 5b. Does the system manage the complete range of digital entities or only a subset?
6. From what precise process(es) or procedure(s), or part thereof, are the digital entities created?
7. To what other digital or non-digital entities are they connected? Is such connection documented or captured?
8. What are the documentary and technological processes or procedures used to identify, retrieve, and access the digital entities?
9. Are those processes and procedures documented? How? In what form?
10. What measures does the creator take to ensure the quality, integrity, and security of the digital entities and their documentation?
11. Does the creator think that the authenticity of his digital entities is at risk?
12. How does the creator use the digital entities under examination?
13. How are changes to the digital entities made and recorded?
14. Do external users have access to the digital entities in question? If so, how, and what kind of uses do they make of the entities?
15. Are there specific job competencies (or responsibilities) with respect to the creation, maintenance, and/or use of the digital entities? If yes, what are they?
16. Are the access rights (to objects and/or systems) connected to the job competence of the responsible person? If yes, what are they?
17. Among its digital entities, which ones does the creator consider to be records and why?
18. Does the creator keep the digital entities that are currently being examined? That is, are these digital entities part of a record keeping system? If so, what are its features?
 - 18a. Do the recordkeeping system(s) (or processes) routinely capture all digital entities within the scope of the activity it covers?
 - 18b. From what applications do the recordkeeping system(s) inherit or capture the digital entities and the related metadata (e.g. email, tracking systems, workflow systems, office systems, databases, etc.)?
 - 18c. Are the digital entities organized in a way that reflects the creation processes? What is the schema, if any, for organizing the digital entities?
 - 18d. Does the recordkeeping system provide ready access to all relevant digital entities and related metadata?
 - 18e. Does the recordkeeping system document all actions/transactions that take place in the system re: the digital entities? If so, what are the metadata captured?
19. How does the creator maintain its digital entities through technological change?
 - 19a. What preservation strategies and/or methods are implemented and how?
 - 19b. Are these strategies or methods determined by the type of digital entities (in a technical sense) or by other criteria? If the latter, what criteria?
20. To what extent do policies, procedures, and standards currently control records creation, maintenance, preservation and use in the context of the creator's activity? Do these policies, procedures, and standards need to be modified or augmented?
21. What legal, moral (e.g. control over artistic expression) or ethical obligations, concerns or issues exist regarding the creation, maintenance, preservation and use of the records in the context of the creator's activity?
22. What descriptive or other metadata schema or standards are currently being used in the creation, maintenance, use and preservation of the recordkeeping system or environment being studied?
23. What is the source of these descriptive or other metadata schema or standards (institutional convention, professional body, international standard, individual practice, etc.)?

Archive a Digital Twin

- Test the preservation of Carleton's Digital Twin
- Carleton University Archives

Digital Twin

A digital twin is

- An ecosystem of multi-dimensional and interoperable subsystems of
 - things in the real world, natural & human made
 - a digital replica of those things
 - a system that communicates between the digital and the real-world things
 - and the people and institutions that govern, contribute to, use and share it
- Primarily w/in architecture, engineering, construction, and operations (AECO)

Digital Twin Application Areas

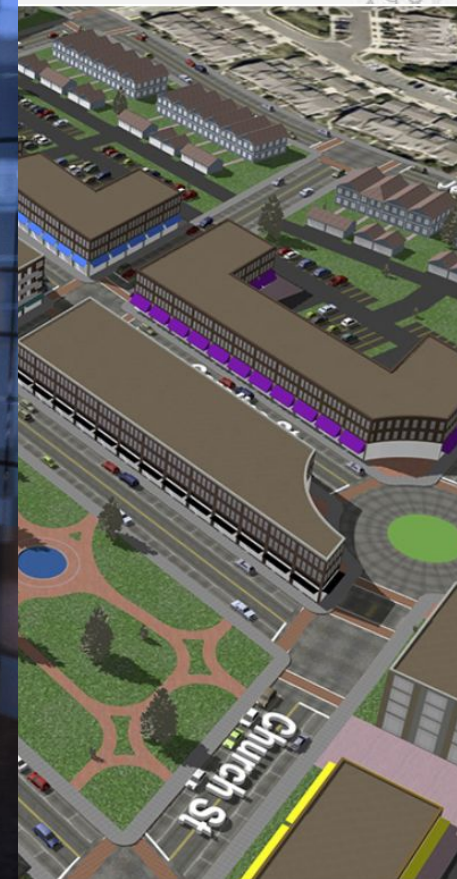
- **Physically large projects**
 - Buildings, bridges & complex structures bound by strict rules of engineering.
- **Mechanically complex projects**
 - Jet turbines, automobiles and aircraft to help improve efficiency w/in complicated machinery & mammoth engines.
- **Power equipment**
 - This includes the mechanisms for generating power and transmitting it.
- **Manufacturing projects**
 - Help streamline process efficiency, as in industrial environments w/ co-functioning machine systems.

Digital Twin





Controlling a playable character a between development scenarios / experience. Photo Credit: Natalie



sville to better visualize the io B shown above.) Photo

AI & Digital Twins

“BIM model to semantically segment laser scan point clouds. These segmented point clouds could then in turn be used to train an AI point cloud classification algorithm. CIMS also worked to manually segment a data set of gothic style ‘high heritage’ value rooms to assist the training of an AI point cloud segmentation algorithm, which commonly has troubles with the variations present in heritage spaces”.



CS04 - Digital Twin



- Sustain Digital Campus project conducted experimentation in 4 test buildings at Carleton U.
- 8 BIM buildings at Carleton
- DT ecosystem of technologies & data allows Architecture, Engineering, Construction, and Operations (AECO) professionals analyze real-time data and visualize the modeling and analysis results into their actual designs, improving the overall process and to manage these assets.

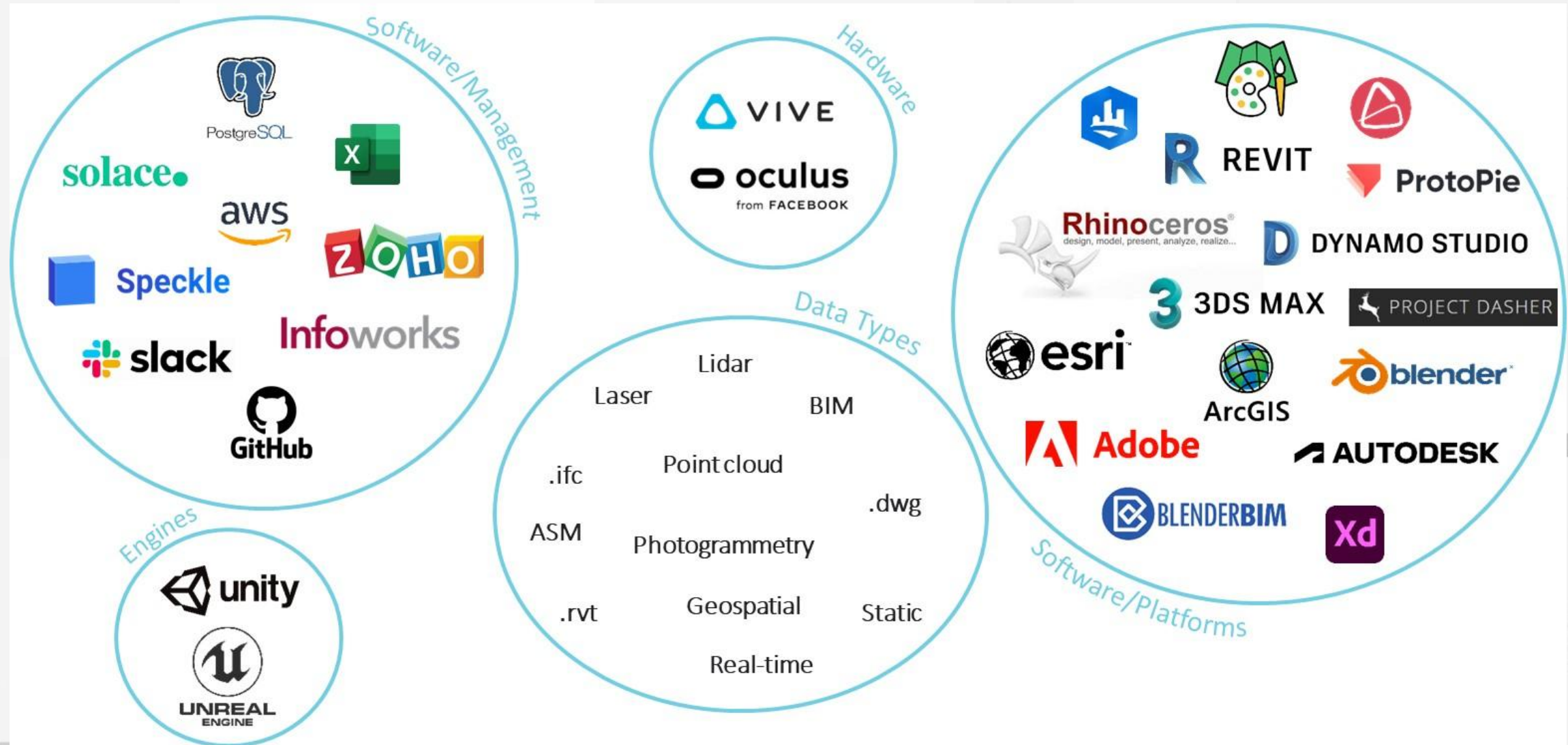
Imagining Canada's Digital Twin

CU04 – Digital Twin Objectives

- A case study to understand how a digital twin is used and created so that it can be preserved
- Test the preservation of the Carleton University Digital Twin created as part of the digital campus work at the Carleton Immersive Media Studio (CIMS) and the Sustain Project
- By doing so, is it possible to use AI to automate the preservation of digital twins and related technologies
- How can DT AI/ML + IoT be preserved



Sample of CIMS DT Data & Technology



CS04 – Digital Twin

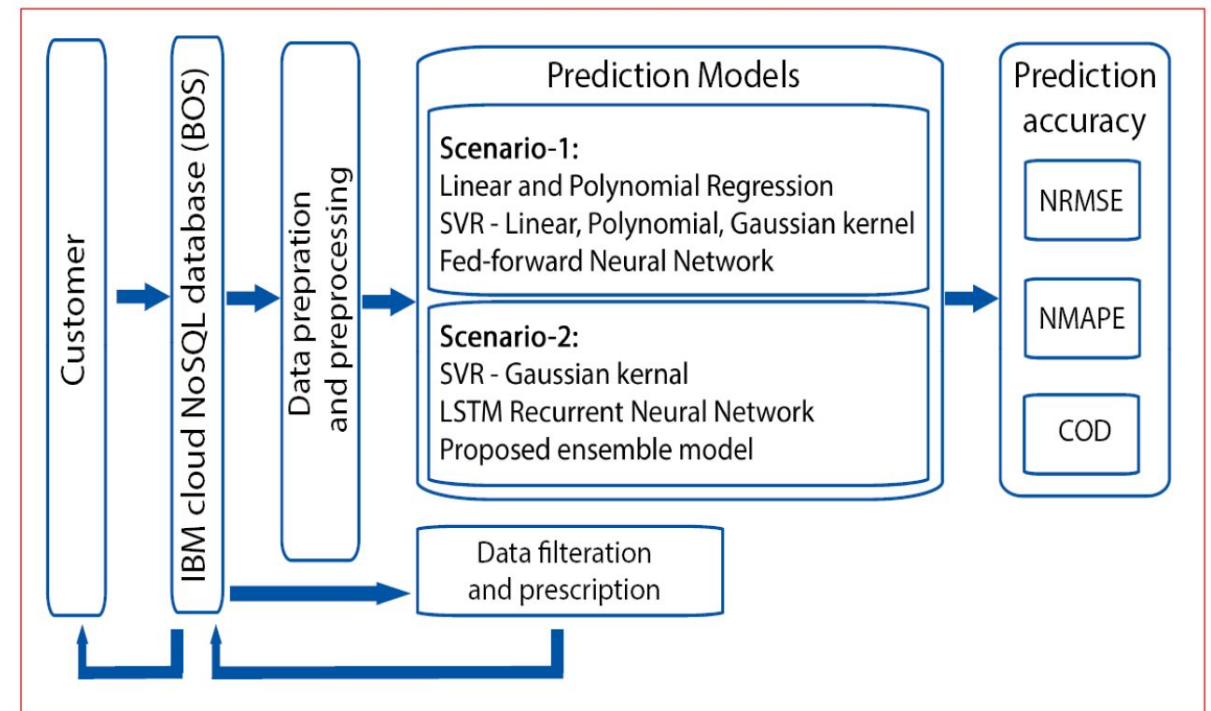
1. Provides critical records and archival challenges in terms of the use and creation of complex records – BIM, ASM, Digital Twins, VR and AI
2. Involves and ecosystem of data, real-time systems, AI/ML, IoT, building information systems (BIM), asset management systems (AMS), LIDAR, VR, simulations and techniques, technologies, and data of importance to fields in architecture, urbanism, planning, construction and engineering.
3. Archival concepts of authenticity, reliability, and accuracy, and others will be identified as research progresses.



CS03 - Smart Grid Data Communication and Analytics

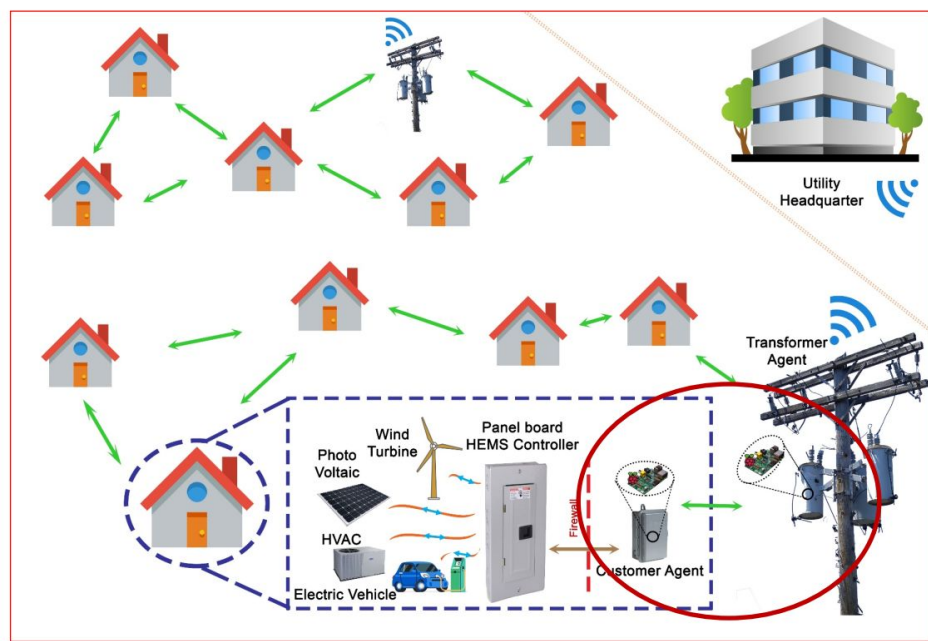
- Smart grids monitor energy consumption and manage production and distribution using a variety of AI/ML and algorithms to predict consumption to manage the grid at power generation but also at the individual household with billing and consumption information
- Common a central element in most smart city strategies, smart grids must address archival and record keeping concerns in order to support accountability of decision-makers and transparency to clients/consumers

Prediction Engine



Sensor Systems and The Internet of Things LAB

CS03 - Smart Grid Case Study



- The Sensor Systems and The Internet of Things Lab at Carleton University conducts research with the Ontario Smart Grid that uses AI/ML in its IoT System
- To inform how archival concepts and principles can influence the development of responsible AI; and will enable outcomes to be validated, thus directly addressing objectives:
- Will be a source of information for other studies in connection with explainable AI, creation, use, retention, and preservation.
- Has important implications for utilities worldwide, for instance for The Ontario Smart Grid System

CS03 – Smart Grid Objectives

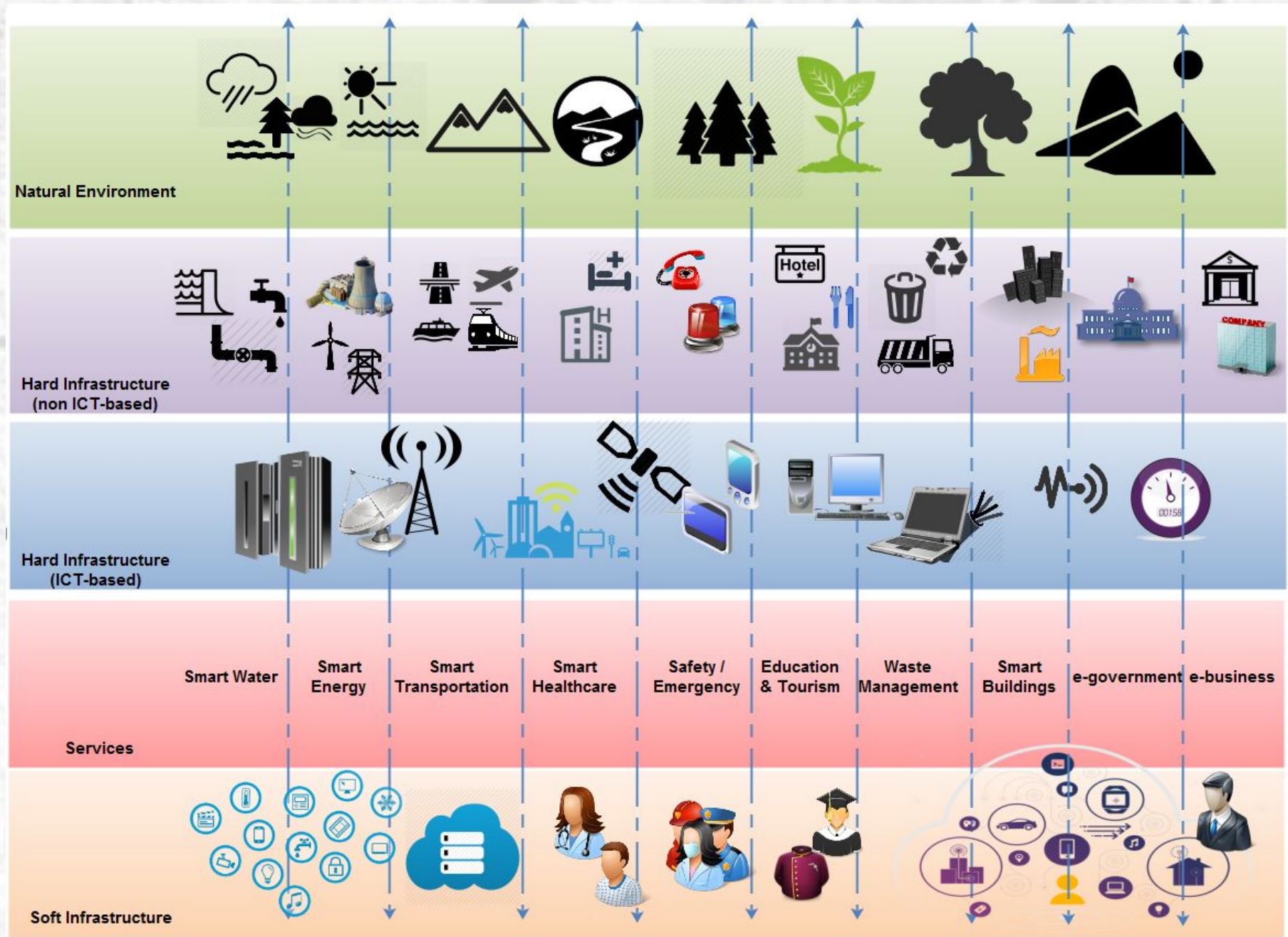
1. The documentation generated in the implementation of a substantial system affecting millions daily by a business of which a democratically elected government and private/public utilities is the majority shareholder, **will provide critical insights into the accountability of a major utility.**
2. Likewise, the findings will clarify the degree to which ML implementations of this scope and complexity are “explainable” to decision-makers and consumers.
3. Good practices identified can be communicated to other jurisdictions and utilities considering similar ML-based enhancements.
4. To conduct an **Analysis of the juridical, provenancial, and procedural contexts of ML-based improvements.**



CS Research have implications for Smart Cities

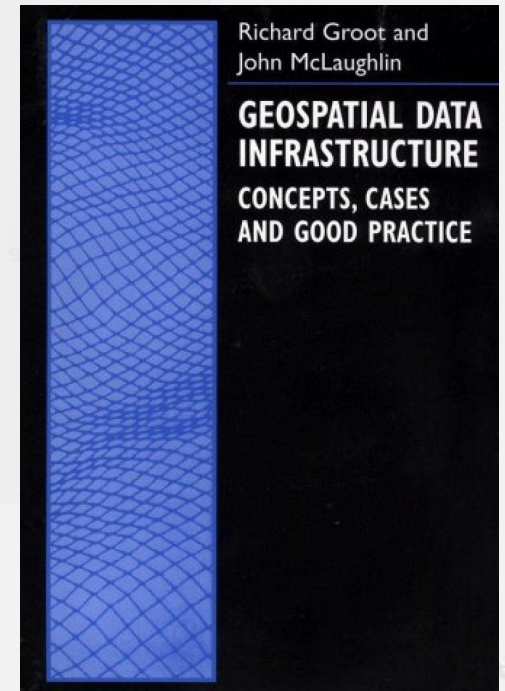
- technologically instrumented & networked w/ systems that are interlinked & integrated, where vast troves of big urban data are being generated by sensors & administrative processes used to manage & control urban life in real-time (Kitchin, 2018).
- where administrators and elected officials invest in smart city technologies & data analytical systems to inform how to innovatively, economically, efficiently & objectively run & manage the city.
- The focus is most often to quantify & manage infrastructure, mobility, business & online government services.

Smart City



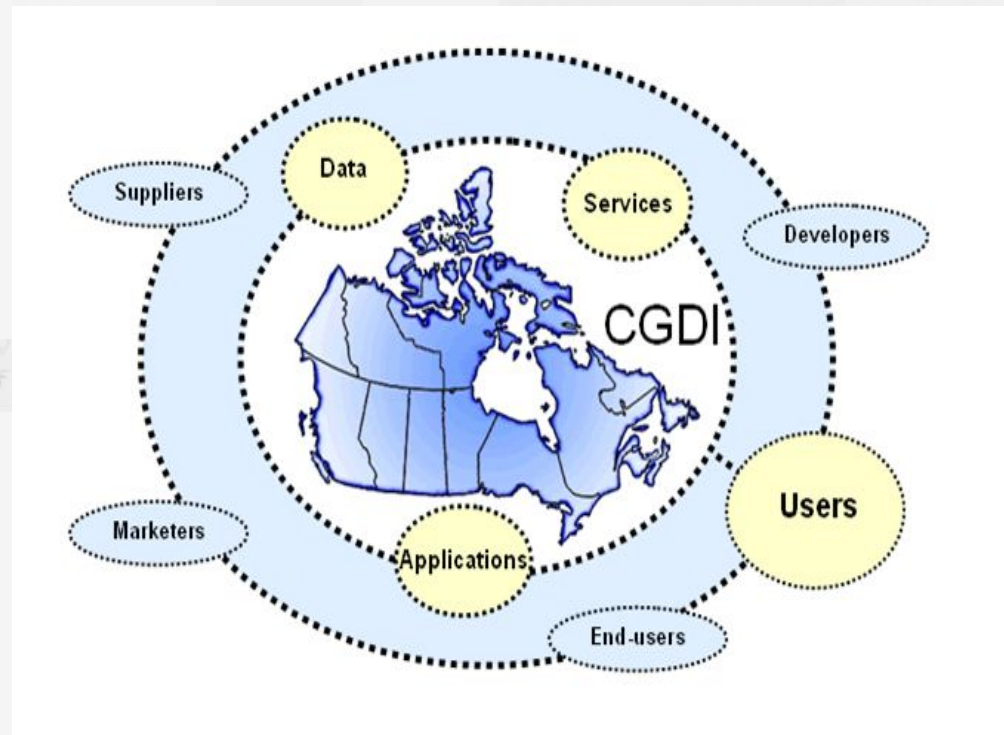
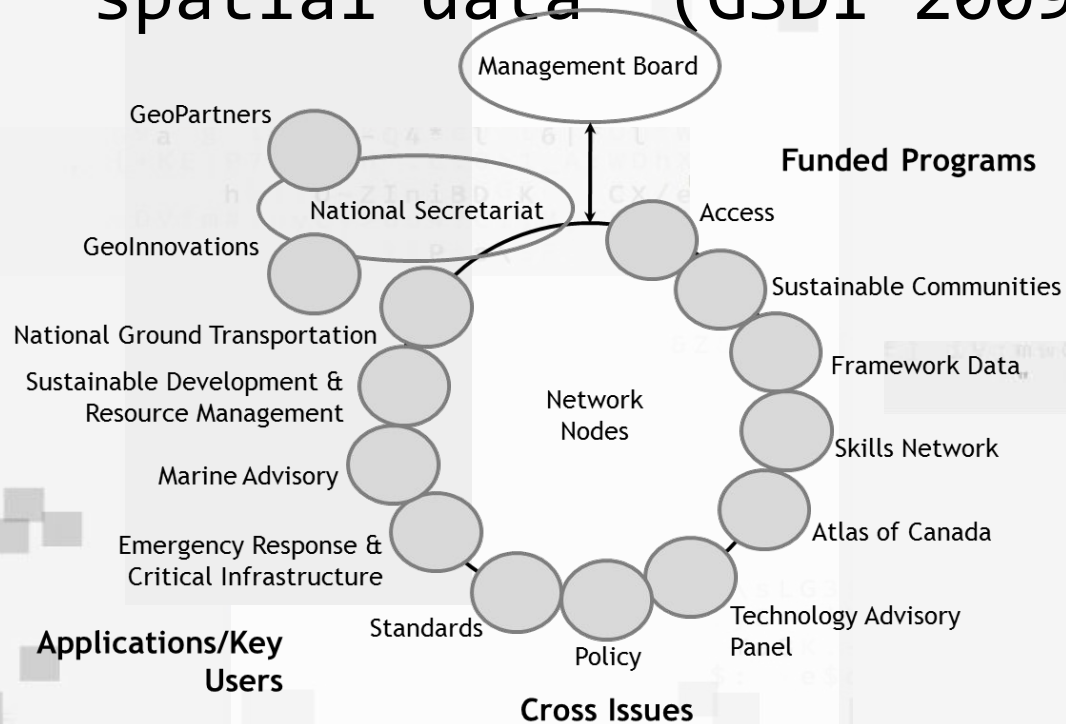
What is a data infrastructure?

- Complex network of institutional, organizational, technological, human and economic resources, which interact with one another to underpin the design, implementation and maintenance of mechanisms that facilitate the sharing, access to, and responsible use of data for a specific application domain or enterprise (Groot & McLaughlin, 2000).



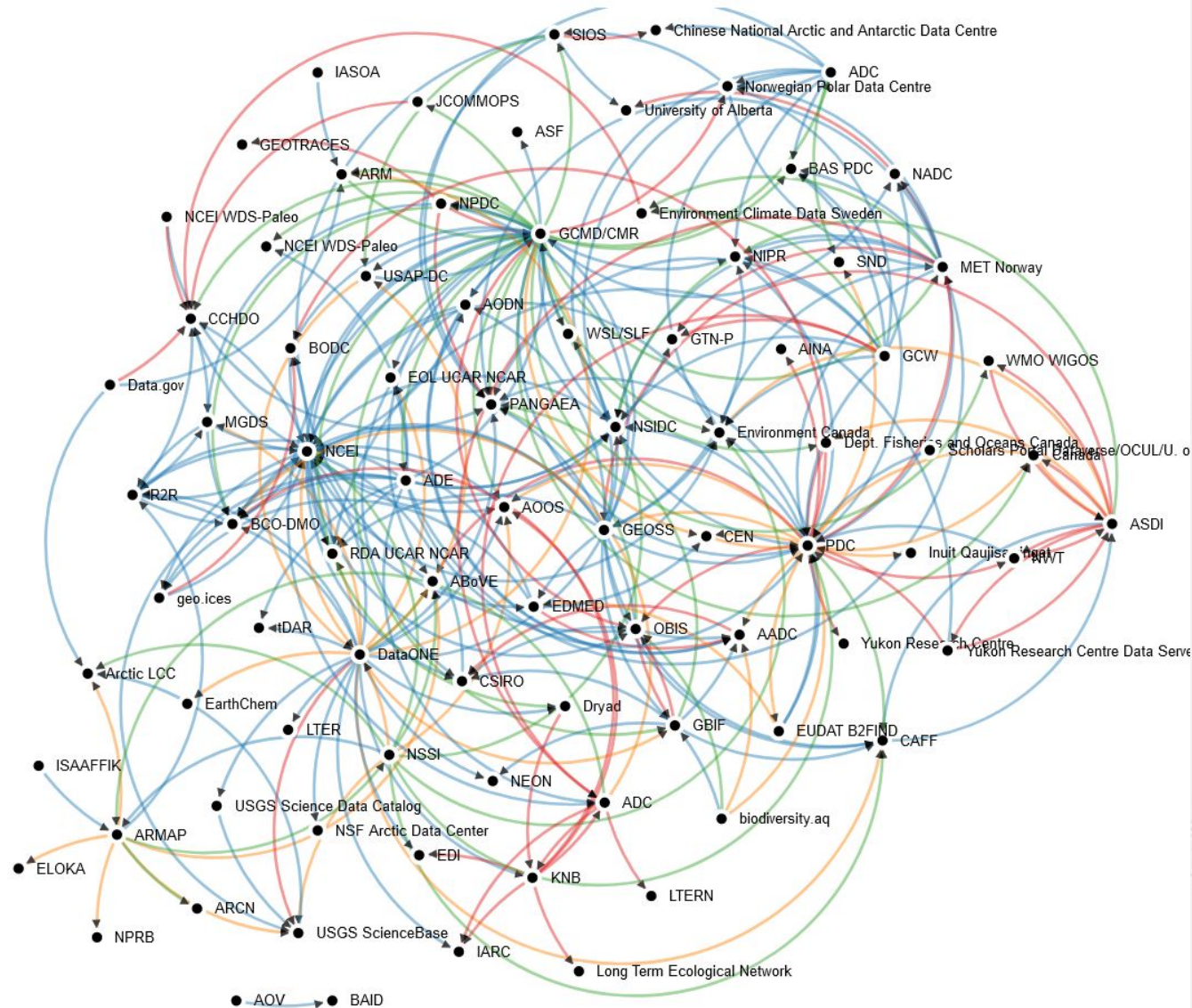
Are related to Spatial Data Infrastructure

- “the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of, and access to, spatial data” (GSDI 2009)



Filter Selection

- Organization
- in Present
- in Past
- Its Complicated
- in Planning



Regional/Polar Data Infrastructure

ASDI Governance



Arctic SDI Services Map Gallery Documents Calendar

Strategic Documents

Who and What is the Arctic SDI?

- 2019-2021 Arctic SDI Biennial Report Highlights from the Icelandic Chairship
- 2017-2019 Biennial Report: Highlights from the Finnish Chairship
- 2015-2017 Biennial Report: Highlights from the US Chairship
- Arctic SDI factsheet September 2018
- Arctic SDI Geportal Factsheet September 2018

Pan-Arctic Digital Elevation Model

- ArcticDEM – Arctic SDI Board Position Statement
- Polar Geospatial Center ArcticDEM Documentation



Governing Documents

- Signed Memorandum of Understanding – English, French, and Russian version
- Signed MoU Extension Document 2019 – English, French and Russian version
- Arctic SDI-Governance_v3.1
- Working Group Membership – 2021
- Arctic SDI Organization Feb 2021

Arctic SDI Strategic Plan Documents

- Strategic Plan 2020-2025
- Roadmap and Implementation Plan 2020-2025

Joint Statement of Intent:

- Arctic SDI – Arctic Regional Hydrographic Commission 2020

Arctic Spatial Data Pilot

- Open Geospatial Consortium Spatial Data Pilot with data intensive scenario based videos and a Final Engineering Report

Arctic SDI Historical Framework

- Arctic-SDI-Framework-Documents_V2.0
 - Provides historical framework and technical vision for the Arctic SDI



Arctic Council stakeholder dialogue and development of the Arctic SDI

- Dialogue with Arctic Council stakeholders identifying needs, requirements and possible contributions
- Develop the Arctic Spatial Data Infrastructure and its map and metadata services providing access to additional reference geodata and thematic data
- Pursue Open Data standards, emerging technologies and industry best practices to remain relevant and interoperable
- Develop governance and guidelines on standards, technical components and services.

2018 News

- Committed engagement in the development of an Arctic Digital Elevation Model (Arctic DEM)
- Initiated cooperation with the Arctic Regional Hydrographic Commission's Arctic Marine SDI Working Group to facilitate access to Arctic marine data and integrate sea and land data

The Arctic SDI cooperation

The Arctic SDI governance model is based on cooperation on prioritized activities where, as agreed to in the Arctic SDI Memorandum of Understanding, activities are developed and managed through the voluntary commitment of each agency. You can learn more about the Arctic SDI at the website: arctic-sdi.org/.

History of the Arctic SDI

The Arctic SDI concept was introduced in 2007 and the Arctic Council Senior Arctic Officials unanimously gave formal support to the Arctic SDI initiative in 2009. The signing of a Memorandum of Understanding (MOU) in 2014 led to demonstrable progress toward building the first elements of the Arctic SDI and approving a new governance model as well as the Strategic Plan 2015-2020.



The role of the 8 National Mapping Agencies of the Arctic countries

- Provide open access to a coherent and authoritative Arctic reference map and thematic Arctic data through the publication of selected data from their respective holdings and from other sources
- Lead and guide the development of an Arctic Spatial Data Infrastructure to further international SDI best practices.

Challenges

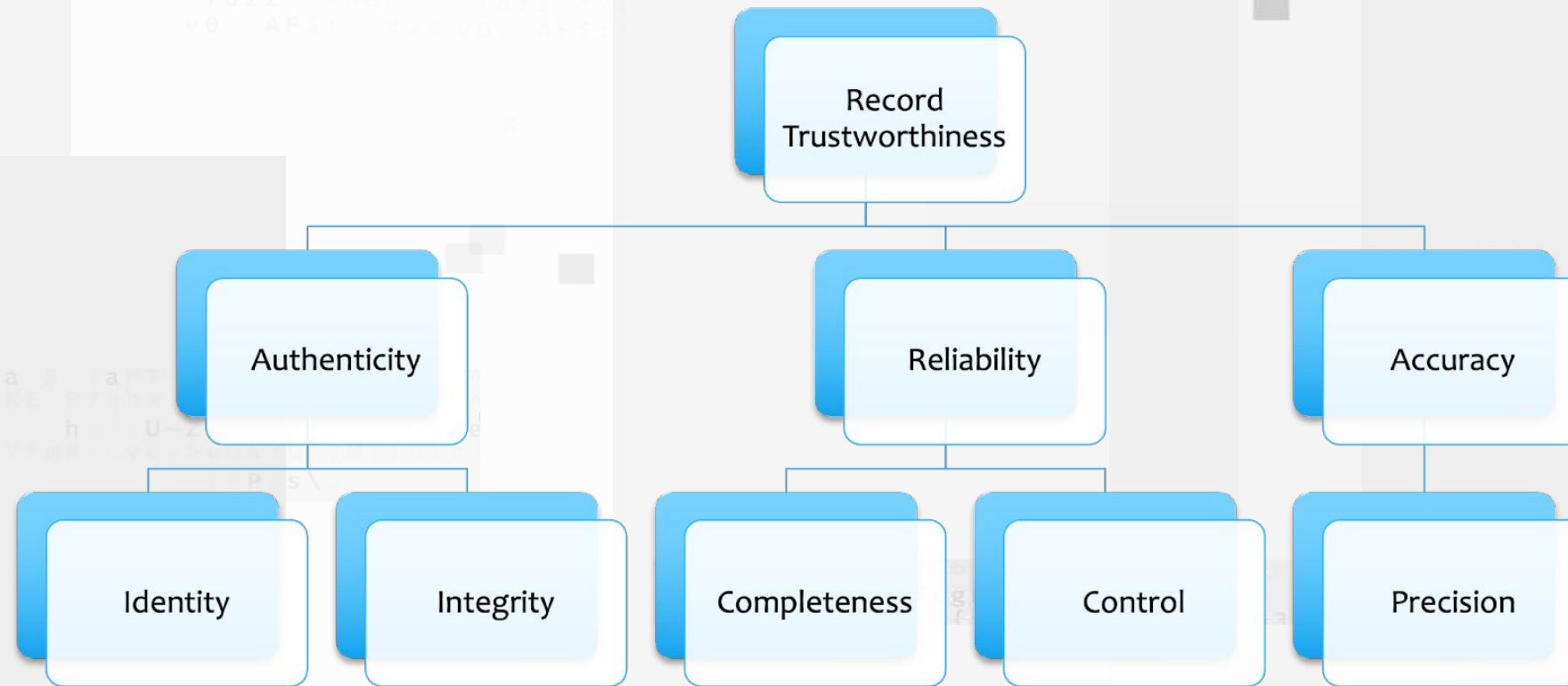
Scale

- **Digital Twin**
 - Emergent
 - Maybe distributed or centralized
 - Not yet standardized
 - ACEO, cities, manufacturing
- **Smart City**
 - Establishing
 - Centralized
 - Standardizing
 - Corporate & city actors
- **Spatial Data Infrastructure**
 - Established
 - Federated & Centralized
 - Standardized & Interoperable
 - Government, research and corporate

- **Commonalities**

- Complex data & technological ecosystems
- Heterogeneous data
 - Real time data
 - Big & small data
 - IoT
 - Framework
- Many platforms
- Standards
- Multiple types of AI at work
- Scale shifting
- Not very standardized
- Many actors

Assessing Trustworthiness in a complex social and technical system?



Duranti, Rogers, Sullivan, 2022

InterPARES Trust

Structure of records in a complex ecosystem?

Medium	Physical carrier of the content
Form	Rules of representation, both physical and intellectual
Content	Message
Act	Reason the record exists
Persons	Agents participating in the action/record
Archival Bond	Relationship of the record to other records participating in the same action
Context	Framework of contexts from general to specific in which the record is created and preserved
Form, Function, Structure, Elements of Form, Protocols...	

Duranti, Rogers, Sullivan, 2022



Challenge of Archival Studies meets Critical Data Studies

Transdisciplinary Research

- “is defined as research efforts conducted by investigators from different disciplines working jointly to create new conceptual, theoretical, methodological, and translational innovations that integrate and move beyond discipline-specific approaches to address a common problem”

Harvard Transdisciplinary Research in Energetics and Cancer Center, 2022

Q & A

Projects

InterPARES Trust AI, UBC Canada

- Funded by the Social Sciences and Humanities Research Council of Canada, <https://interparestrustai.org/trust>, @itrustai

Imagining Canada's Digital Twin,

- Carleton Immersive Media Studio (CIMS), Carleton University, Canada
- funded by the New Frontiers in Research Fund (NFRF), <https://canadasdigitaltwin.ca>

Sustain Designing for People,

- Carleton University, Canada
- funded by NSERC, <https://sustain.sce.carleton.ca/>

Sensor Systems and Internet of Things Lab,

- Department of Systems and Computer Engineering, Carleton University, Canada
- Funded by CISCO and others <https://carleton.ca/internetofthings/people/dr-mohamed-ibnkahla/>

Image Source CS04 & CS06

CS03 & CS04 Images come from:

- Cover Slide - CIMS
http://cims.carleton.ca/#/projects/imagining_canada's_digital_twin
- [Sensor Systems and The Internet of Things Lab](#) presentation at the WG1 Meeting on October 25, 2021 by Zied Bouida.
- [Carleton Immersive Media Studio](#) (CIMS) and the presentation at the WG1 Meeting on October 25, 2021 by Nicolas Arellano
- **Contact:**
 - Tracey.lauriault@carleton.ca
 - @TraceyLauriault

CS04 – Digital Twin Research Team

- (I Trust AI Lead) **Tracey P. Lauriault**, Associate Professor, Critical Media and Big Data, School of Journalism and Communication, Carleton University (Tracey.Lauriault@Carleton.ca)
- (I Trust AI Co-Lead) **Sherry Xie**, Renmin University of China - School of Information Resource Management (sherry1x@outlook.com)
- **Stephen Fai**, Professor cross appointment in the Department of Civil and Environmental Engineering, the Institute for Comparative Studies in Literature, Art, and Culture, and the Azrieli School of Architecture and Urbanism; Director of the Carleton Interactive Media Studio (CIMS) (sfai@cims.carleton.ca)
 - **Nicolas Arellano**, Ph.D. Candidate at Carleton University | Research Team Lead at CIMS (narellano@cims.carleton.ca)
- **Travis Kinnear** (he/him/his), MLIS, Digital Archivist, Corporate Records and Archives, University Secretariat, Carleton University (Travis.Kinnear@carleton.ca)
- **Tero Päivärinta**, Professor, Luleå University of Technology, Research subject: Information systems, Digital Services and Systems, Department of Computer Science, Electrical and Space Engineering, (tero.paivarinta@ltu.se)
- **Göran Samuelsson**, Senior lecturer, Department: Department of Information Systems and Technology (IST) and Information Specialists for The Swedish Transport Administration (Goran.Samuelsson@miun.se)
goran.samuelsson@trafikverket.se)
- **Liam O'Brien**, Professor in Architectural Conservation and Sustainability Engineering, PI the SUSTAIN Project (Sensor-based Unified Simulation Techniques for Advanced In-Building Networks) project at Carleton University (liam.obrien@carleton.ca)
- **H. Burak Gunay**, Assistant Professor in Building Science, Department of Civil and Environmental Engineering (burak.gunay@carleton.ca)
- **Annantonia Martorano**, Dipartimento di Storia, Archeologia, Geografia, Arte e Spettacolo - SAGAS, University of Florence (UNIFI) (annantonia.martorano@unifi.it)
 - **Elena Gonnelli**, PhD student at the University of Florence., (gonnellielena@gmail.com elena.gonnelli@unifi.it)
 - **UNIFI research Assistants**
 - Raquel Jennyfer De Souza raquel.desouzasantos@stud.unifi.it;
 - Luca Tondini, Irene Nocentini, Erica Cavina, Valentina Vazzano

CU04 – Digital Twin

Lead:

- ***Tracey P. Lauriault***, Carleton U.
- (Co-Lead) ***Sherry Xie***, Electronic Records Management Research Center, Renmin U. of China

Team:

- ***Stephen Fai***, Carleton Interactive Media Studio (CIMS)
- ***Nicolas Arellano***, Ph.D. Candidate, Carleton U.
- ***Travis Kinnear***, Digital Archivist, Carleton U.
- ***Tero Päivärinta***, Luleå University of Technology, Electrical and Space Engineering
- ***Göran Samuelsson***, Swedish Transport Administration
- ***Liam O'Brien***, SUSTAIN Project (Sensor-based Unified Simulation Techniques for Advanced In-Building Networks) Carleton U.
- ***H. Burak Gunay***, Civil and Environmental Engineering, Carleton U.
- ***Annantonia Martorano***, SAGAS, U. of Florence (UNIFI)
- ***Elena Gonnelli***, PhD Candidate, U. of Florence

CS03 – Smart Grid Team

Lead

- **Tracey P. Lauriault**, Associate Professor, Critical Media and Big Data, Communication and Media Studies (Tracey.Lauriault@carleton.ca)

Team

- **Mohamed Ibnkahla**, Professor, CISCO Chair in the Internet of Things & Director Internet of Things Lab, Department of Systems and Computer Engineering (MohamedIbnkahla@cunet.carleton.ca)
- **Zied Bouida**, Ph.D., PEng, SMIEEE, Senior Researcher, Cisco Chair Manager, Internet of Things Lab, Department of Systems and Computer Engineering (Arrangement & Description) (ziedbouida@sce.carleton.ca)
- **Jim Suderman**, retired archivist, considerable archival and program management experience in the public sector (Jim.Suderman@gmail.com)
- **Sanja Seljan**, University of Zagreb, Faculty of Humanities and Social Sciences. (sanja.seljan@FFZG.HR)
- **Chris Trainor**, (he/him), Head, Archives & Special Collections, MacOdrum Library, Carleton University (ChrisTrainor@cunet.carleton.ca)