Big Data Fundamentals and Impact

A WIL Digital Initiative

Presented by the Information and Communications Technology Council

Welcome to ICTC's WIL Digital: Big Data Fundamentals and Impact

ICTC's WIL Digital is an innovative Work Integrated Learning program that helps students gain insight and skills to advance learning and meaningful work experience.

COURSE OVERVIEW

Welcome to "Big Data Fundamentals and Impact"! This ICTC course is a great opportunity for all our participants to learn about one of the most exciting industries today. It also serves as a source of knowledge of big data technologies and their current impact on companies around the world, and especially in Canada.

ABOUT THIS COURSE

As you may notice (based on the file of the course), there are two main learning blocks: big data fundamentals, and legal and ethical impacts in Canada. Each contains three modules for a total of six modules.

Module	Title	Learning goals
1	Value of Big Data and Potential Applications	Learning big data technologies, fundamentals, and its relationship with other technologies such as artificial intelligence.
2	Working with Data	Analyzing big data architectures, data integration methodologies, and companies providing data as a service.
3	Data Teams and Roles	Describing data teams, typical roles, activities, and tools.
4	Business Models and Strategic Analysis	Leveraging existing tools to analyze business models and company strategies and determining big data project value.
5	Legal Implications	Comparing data and AI regulations around the world.
6	Ethical Impacts	Exploring data ethics approaches, guidelines, and principles.
	Case Study	Self-Sovereign Identity (SSI) Systems

We believe that these modules will provide an updated view of the big data industry and its state-of-the-art applications.

We also recommend you read this ICTC-CTIC report on big data and the intelligence economy: https://www.ictc-ctic.ca/wp-content/uploads/2015/12/BIG-DATA_2015.pdf, and the updated information provided in the ICTC-CTIC report on digital talent outlook 2025: https://www.ictc-ctic.ca/wp-content/uploads/2021-08/digital-talent-outlook-for-2025.pdf. It explores the immense big data opportunity for companies and job seekers in the Canadian market. We will refer to specific sections of this document throughout the course.

Targeted audience

This course is designed for students who are interested in developing fundamental skills in the subject of big data as an entry-level introduction to this topic. Students who are already proficient in this subject matter are encouraged to take another of the WIL Digital courses available, which provide an excellent lateral understanding of concepts associated with this course. If you are interested in studying a more in-depth version of this subject, please contact your course facilitator for available options.

Each module is approximately 1-2 hours long, to be completed during your work placement. The suggested timeline to complete is 1 week per module and 4 weeks to complete the case study.

As a participant, you will:

- Have access to 6 unique micro-learning modules, supported by industry partners in the big data ecosystem across Canada
- Find ways to jump start your career in big data
- Add new skills to your professional portfolio
- Become the next generation of talent in the big data landscape!

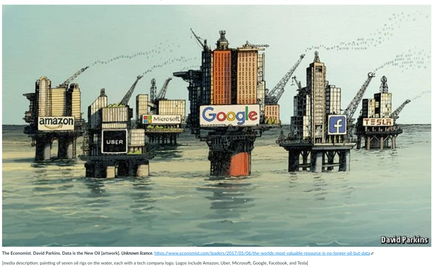
Big Data Fundamentals and Impact

A WIL Digital Initiative

Presented by the Information and Communications Technology Council

1.1 Introduction to Big data, and Examples in Various Industries Around the World

Big data is sometimes considered as a buzzword, but the reality is that this technology (or set of technologies) is already transforming industries and organizations around the world. Data has been considered the "new oil" for years because of its tremendous potential, but also because of the need for refinement to extract actual value. Companies are investing millions to create data teams and to develop new data-driven applications.

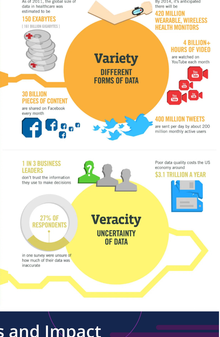


The Economist, **David Parkins**, **Data is the New Oil (artwork)**, **Unknown Ikonos**. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

[media description: painting of seven oil rigs on the water, each with a tech company logo. Logos include Amazon, Uber, Microsoft, Google, Facebook, and Tesla]

But what do we mean by big data? Is there also something called small data? If yes, how to make the difference? Is it only about the size?

Technically, big data is not only "big." This means, we don't consider only the volume (amount or size of data) but also other dimensions such as variety, velocity, veracity, etc. Note: there is a panoply of versions with three, four, five, seven, and even ten "V." Feel free to research the other dimensions, but these four are the most relevant as data "characteristics".



Big Data Fundamentals and Impact

A WIL Digital Initiative

Presented by the Information and Communications Technology Council

2.0 Working with Data

[Media description: Hispanic man, with short brown hair and beard, wearing a grey shirt with a black collar. The background is a solid yellow colour.]

2.3 Data collection methods and APIs

Mohammed Hassan. Build develop web site design [Image]. Pixabay licence
https://pixabay.com/illustrations/build-develop-site-web-design-3916373/

Transcript: Welcome back to class. I hope you have enjoyed the first model, model zero and one. If you're here, that means you have already completed the activities, you have completed, but on which you have about your thoughts on module one. I'm very happy to have you here. Today we are going deeper, we'll be bringing in different nations like data architecture. We'll be talking about the concept of open data, we'll be talking about data as a business. You know, there are some companies making good money with data, enabling it as an API. You will see what that means. But just to say, that this is a very important module, I want you to focus, to pay attention to what we will be doing. Again, we have a combination of text information, different podcasts, etc., etc. I want you to complete everything. Also, check the additional resources. These two models, model one and two together, like the learning block, let's say the big data fundamental, it's very important that you get the value from all of these and you focus on what you are learning on these two models. OK? And after that, we'll be talking about teams and positioning. But I want you to focus on this one for now. So, enjoy it. There's a lot of material, and talk to you later.

Big Data Fundamentals and Impact

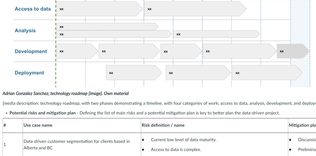
A WIL Digital Initiative

Presented by the Information and Communications Technology Council

4.3 Project management

There is no standard for data project management practices (yet), but this aspect is becoming increasingly important for Canadian (and international) organizations. There are several aspects to be considered in terms of project management for data-driven projects:

- **General data project roadmap** - A roadmap is a mix of a timeline, categories of work, task owners, etc. The idea is to have a time-based plan for the data-driven project. Here is a visual example of what a roadmap would look like (*illustrative example; details are not relevant*):



Adrian Gonzalez Sanchez, technology roadmap [image]. Own material

[media description: technology roadmap, with two phases demonstrating a timeline, with four categories of work; access to data, analysis, development, and deployment.]

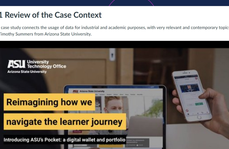
- **Potential risks and mitigation plan** - Defining the list of main risks and a potential mitigation plan is key to better plan the data-driven project.

#	Use case name	Risk definition / name	Mitigation plan
1	Data-driven customer segmentation for clients based in Alberta and BC	• Current low level of data maturity. • Access to data is complex.	• Discussion on data systems with the engineering team for improvements. • Preliminary meeting with involved stakeholders to design a plan for access to data.
2			
...			

- Risk name: e.g., potential delays during the second part of the project
- Mitigation plan: e.g., adjustment of project scope or available resources

- **Project phases** - How to structure a data-centric project end-to-end
  - Exploration - Discussing potential best projects, depending on the available resources
  - Mobilization - Preparing the required people, data, and infrastructure before starting the execution phase project
  - Execution - Implementing the data transformation, cleansing, etc. and generating results (this will depend on the nature of the project)
  - Delivery - Presenting project results, including data insights
  - Maintenance - Technical post-delivery maintenance (e.g. bugs, improvements, etc.)

- **Project management skills** - Data-centric project management is similar to traditional project management, but it requires special control of the details and project phases. For that to happen, there is a series of activities and related skills which are required, and which are part of the A.D.R.I.A.N. Framework:



Adrian Gonzalez Sanchez. A.D.R.I.A.N.'s framework [Image/s]. CC BY 4.0

[media description: six circles, each with a letter spelling 'ADRIAN'. A is for analyzing, d is for defining, R is for resourcing, I is for involving, A is for agreeing, N is for Navigating/Nurturing.]

These activities are very typical when we are planning and executing data projects. Concretely, there is someone (typically a data project manager) who will take care of the details, guarantee the project progress, and work with the different stakeholders. Main activities are:

- Analyzing requirements, potential risks, project progress, results, etc.
- Defining project phases, involved stakeholders, deliverables, etc.
- Resourcing the project, which includes people, data, infrastructure, etc.
- Involving all project stakeholders, from executive sponsors to the technical team
- Agreeing or enabling agreements, to attain the buy-in from the different project stakeholders

Navigating the obstacles and nurturing the project and company data-driven culture

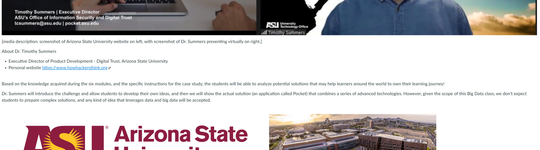
Big Data Fundamentals and Impact

A WIL Digital Initiative

Presented by the Information and Communications Technology Council

7.1 Review of the Case Context

This case study connects the usage of data for industrial and academic purposes, with very relevant and contemporary topics, such as self-sovereign identity (SSI) and self-owned data. Concretely, the case is presented by both the course instructor Mr. Adrian Gonzalez and our SME, Dr. Timothy Summers from Arizona State University.



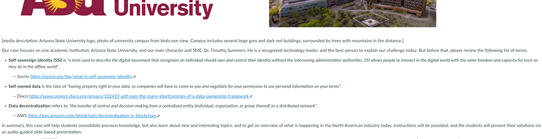
[media description: screenshot of Arizona State University website on left, with screenshot of Dr. Summers presenting virtually on right.]

About Dr. Timothy Summers

- Executive Director of Product Development - Digital Trust, Arizona State University
- Personal website: https://www.timothysummers.net

Based on the knowledge acquired during the six modules, and the specific instructions for the case study, the students will be able to analyze potential solutions that may help learners around the world to own their learning journey!

Dr. Summers will introduce the challenge and allow students to develop their own ideas, and then we will show the actual solution (an application called Pocket) that combines a series of advanced technologies. However, given the scope of this Big Data class, we don't expect students to prepare complex solutions, and any kind of idea that leverages data and big data will be accepted.



[media description: Arizona State University logo, photo of university campus from birds-eye view. Campus includes several large grey and dark red buildings, surrounded by trees with mountains in the distance.]

Our case focuses on one academic institution, Arizona State University, and our main character and SME, Dr. Timothy Summers. He is a recognized technology leader, and the best person to explain our challenge today. But before that, please review the following list of terms:

- **Self-sovereign identity (SSI)** - *"a term used to describe the digital movement that recognizes an individual should own and control their identity without the intervening administrative authorities. SSI allows people to interact in the digital world with the same freedom and capacity for trust as they do in the offline world."*
  - Source: https://auth.ca/sovereign-has-what-is-self-sovereign-identity
- **Self-owned data** is the idea of *"having property right in your data, so companies will have to come to you and negotiate for your permission to use personal information on your terms".*
  - Dilos: https://www.project-disco.org/privacy/102619-self-own-the-many-shortcomings-of-a-data-ownership-framework
- **Data decentralization** refers to *"the transfer of control and decision-making from a centralized entity (individual, organization, or group thereof) to a distributed network".*
  - AIRA: https://aira.amazon.com/blog/decentralization-in-blockchain

In summary, this case will help students consolidate previous knowledge, but also learn about new and interesting topics, and to get an overview of what is happening in the North American industry today. Instructions will be provided, and the students will present their solutions via an audio-guided slide-based presentation.

Funded by the Government of Canada | Canada