Identifying cross-cutting capabilities to extract information from big geospatial datasets

Birds of a Feather - C3DIS 2021
Wednesday 7 July

Geospatial Capabilities Community of Practice

We acknowledge and celebrate the First Australians on whose traditional lands we meet, and we pay our respect to the elders past, present and emerging.

Geospatial Capabilities Community of Practice

Established in September 2019

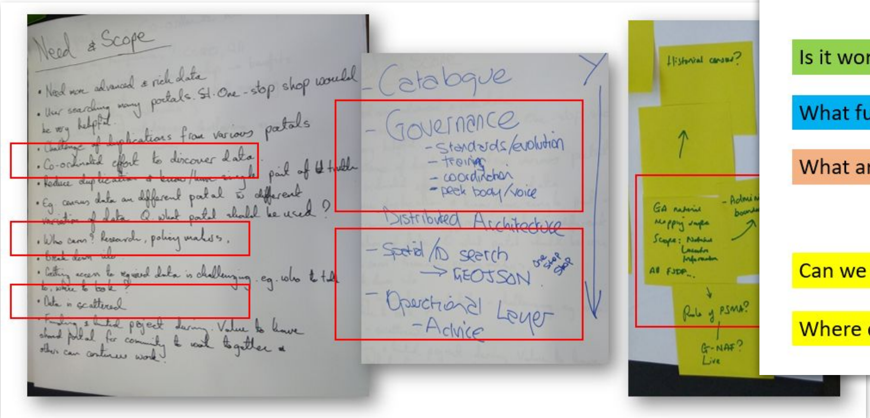
Purpose

- bring people together as a community who, in some capacity, work with geospatial data and make it available for broader use
- a forum for sharing knowledge and experiences with finding, accessing, transforming, using and making available geospatial data

Meetings are held online and occur every couple of months.

https://sites.google.com/ardc.edu.au/geospatialcapcop

eResearch 2019 BoF

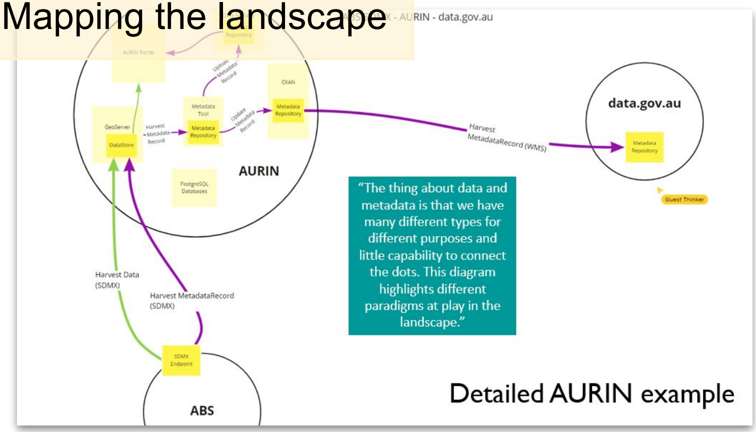


eResearch 2020 BoF

Questions

- Is it worthwhile to capture and represent services in a register?
- What functionality would people expect?
- What are the use cases for such a thing?
- Can we coordinate activity nationally?
- Where do we go from here?

Mapping the landscape



Topics we've covered or discussed:

- ✓ Loc-I
- ✓ ANZ Metadata Working Group
- ✓ Elevation Information System (ELVIS)
- ✓ OSGeo Oceania community
- ✓ GDA2020
- ✓ Geoprivacy
- ✓ Communicating the Value of Space and Spatial technologies

Agenda for today

Introduction

Lightning talks: (3 x 5min)

1. Big Geospatial Data: challenges and opportunities (Sanjeev Srivastava)
2. Use of ML in geospatial data (Hassan Talebi)
3. Digital Twins - Gemini principles (Michael Rigby)

Breakout Sessions (25min)

1. What are the key challenges with big geospatial data? (Sanjeev Srivastava)
2. What are the key challenges in the use of AI in Geospatial Data? (Jens Klump)
3. What are the key challenges to consider when incorporating built/natural environment and social data within Digital Twins? (Michael Rigby)

Report back plenary (3min for each group)

Wrap up and future directions

Session notes at https://bit.ly/C3DISGEO

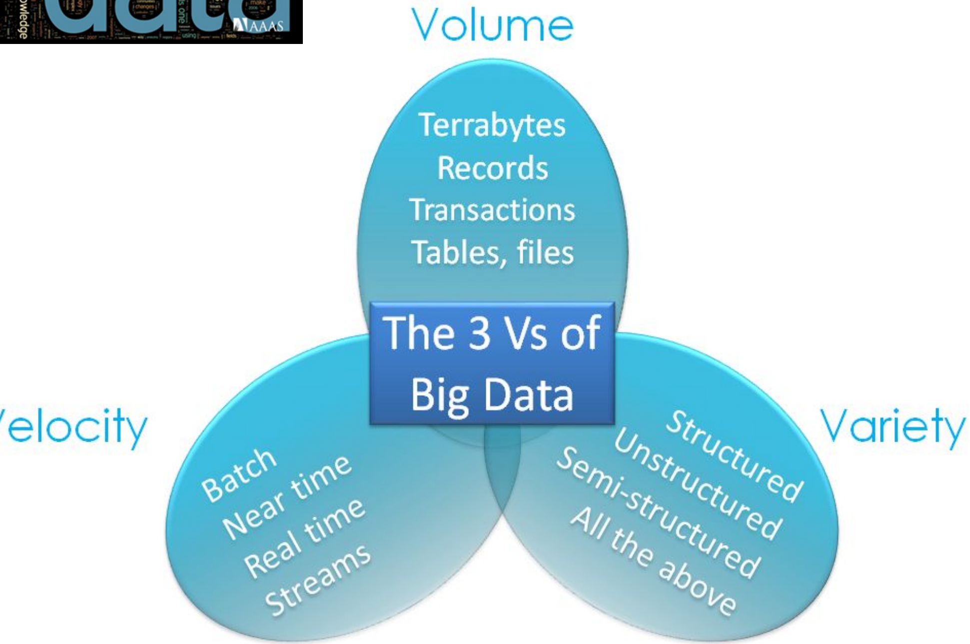
Please register your attendance

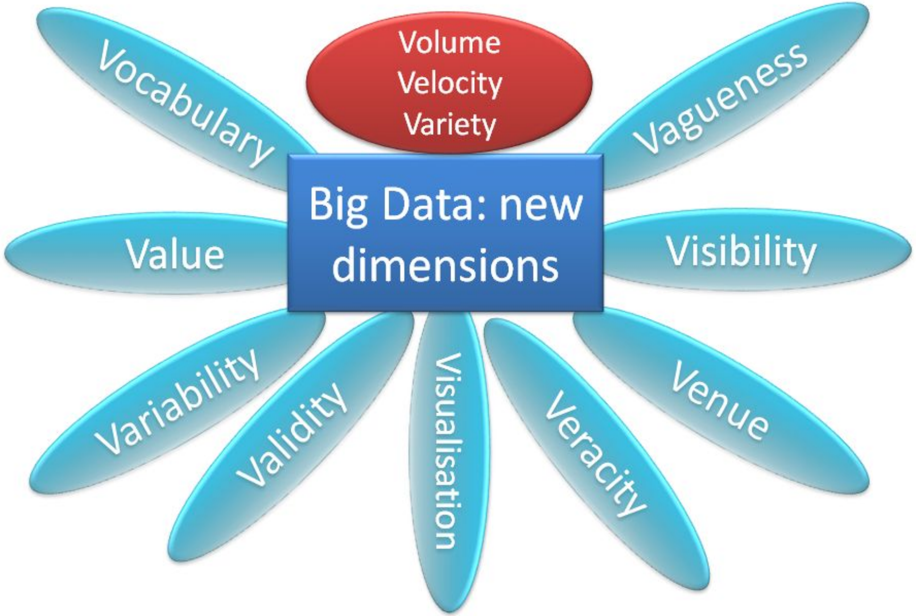
so we know who came

Big geospatial data: challenges and opportunities

Sanjeev Srivastava

What is big data





Big geospatial data types

- Archived geographic information
- Geo-localised data
    - Collected as point
        - using GPS or GPS chip embedded in devices such as mobile phones
        - Using address geocoding
- Spatially grounded
    - location, shape, size, orientation and spatial relationships are integral to the data
        - Remote sensing images (lidar, radar, optical images from satellites and drones).

- Human-sourced
    - Volunteered GIS
        - www.openstreetmap.org
    - Social media
        - Twitter, Flickr

- Process-mediated
    - Government data
        - ABS, BOM
- Consumer data
    - Transactions, loyalty cards etc.
- Machine generated
    - Smart cities sensors (CCTV, traffic flow, etc.)

Big geospatial data formats

Software specific raster/vector/tabular data

Geographic databases (geodatabase, OGC Geopackage)

Point cloud (las)

Web files

Complex scientific data (NetCDF, HDF, GRIB)

Application examples

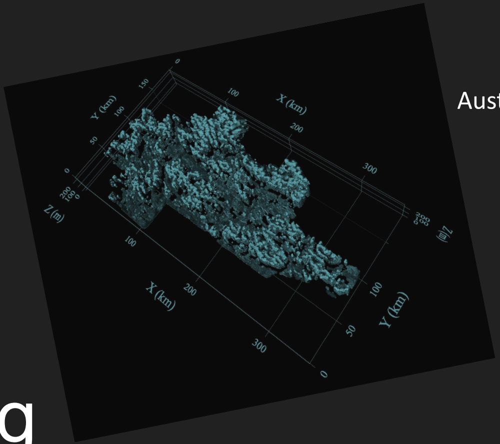
- Built environment
  - Smart cities
  - IoT
  - Mobility
  - Transport
- Natural resources
  - Forest inventory
  - Climatic change
  - Animal tracking
- Health and wellbeing
  - Sentiment analysis
  - Pandemic and endemic hot spots
  - Migration

- Google's flu map

Big geospatial data and geospatial concepts

- Place, space and time
  - Location
  - Context
  - Geographic extent
  - Temporal extent
- Representation
  - Data models
  - Coordinate system
  - Structured/unstructured
  - Generalisation
- Scale
  - Spatial dependance
- Quality
  - Biasness
    - Sample and population
- Users and usability
  - Access
  - Ethics
  - Interaction
- Interoperability
  - Device
  - software/apps
  - Fusion
- Visualisation
  - 2d patterns
  - Digital twin
- Analytics
  - Patterns, coordinated view, user-centred

Australia's National Science Agency



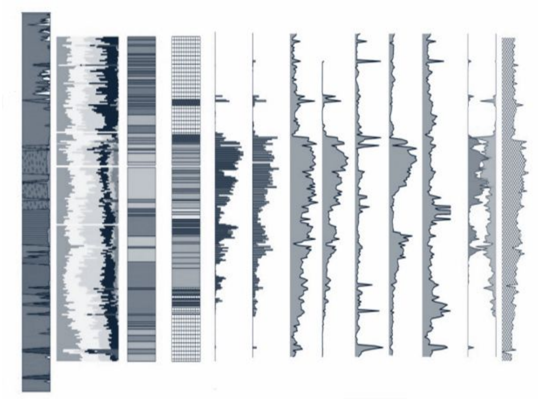
Use of ML in geospatial data: Towards geostatistical learning

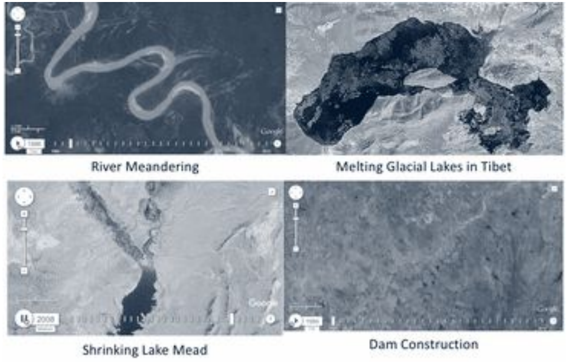
Hassan Talebi | July 2021

CSIRO Deep Earth Imaging FSP, Australia, Hassan.Talebi@csiro.au
CSIRO Mineral Resources, Australia

The particularities of geosystems and geoscience data

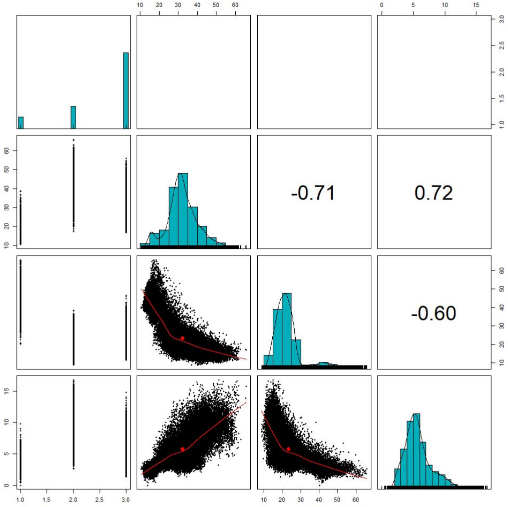
- Big data
- Multi-source and mixed-type
- Multi-scale
- High-dimensionality
- Limited sample size
- Paucity of ground truth information
- Importance of extreme cases
- Spatial and temporal heterogeneity
- Auto- and cross-correlations





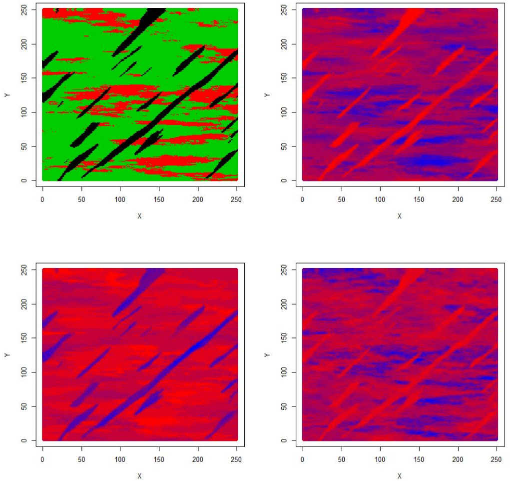
Statistical learning

Nonspatial learners

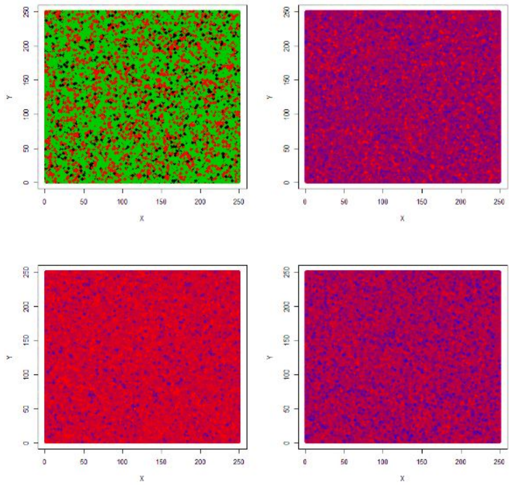


$$RF\#1 = RF\#2$$
$$SVM\#1 = SVM\#2$$
$$LR\#1 = LR\#2$$
$$\ldots = \ldots$$

Case #1

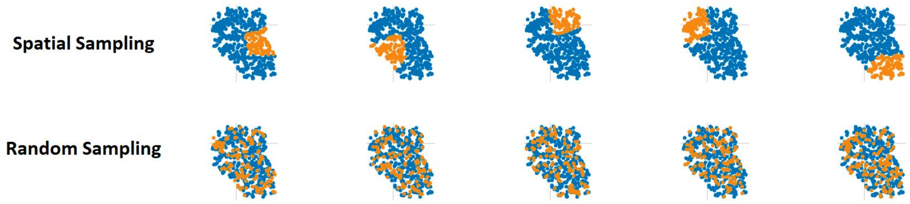


Case #2

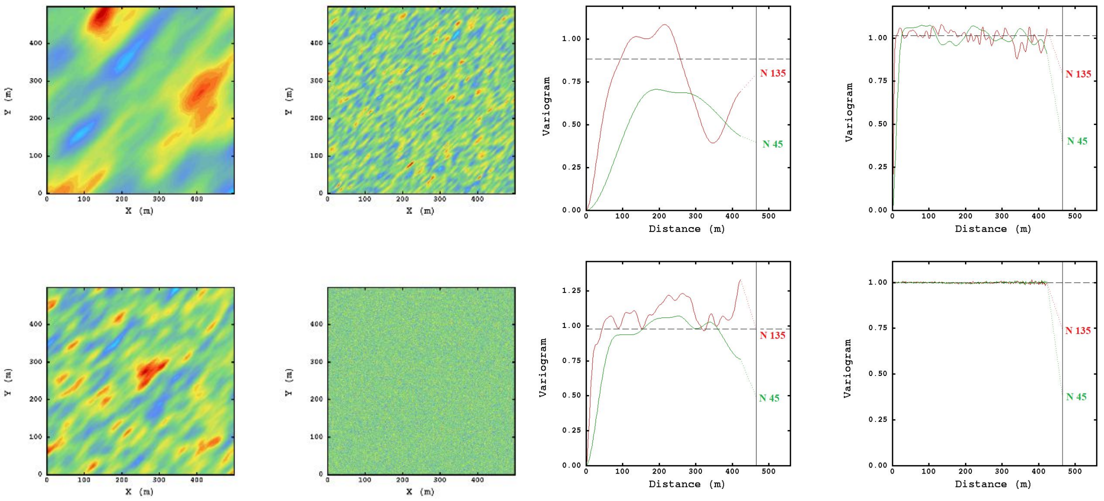


Statistical learning

Nonspatial learners



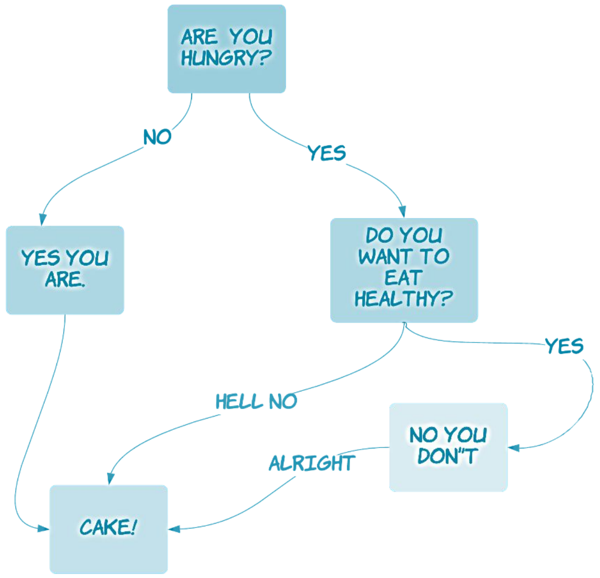
- Independent and identically distributed random variables
- Random sampling and overoptimistic learners



Statistical learning

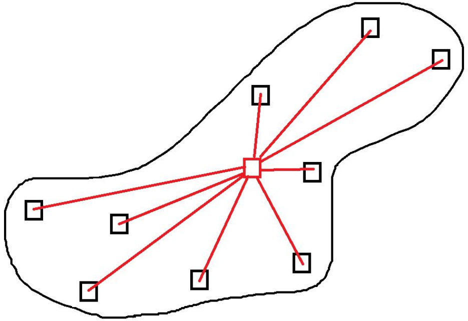
Nonspatial learners

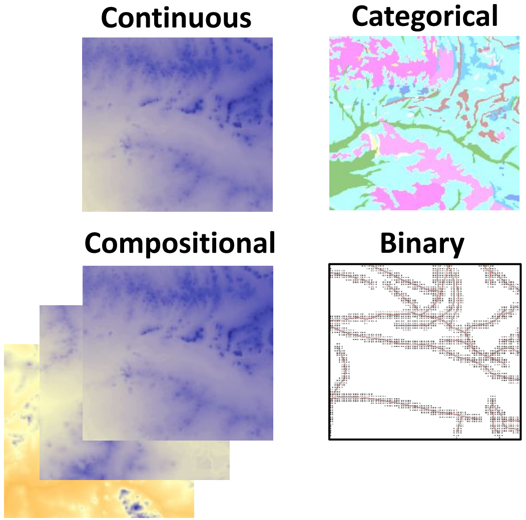
- Incorporating prior physical knowledge in the learning process
- Discovering the physics behind the geological phenomena by machine learning

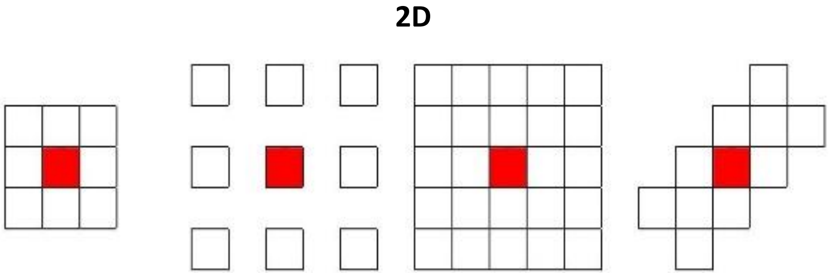


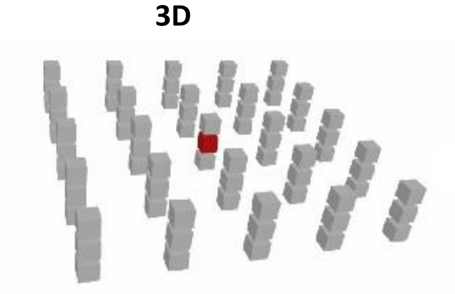
Geostatistical Learning

Improve spatial awareness using higher order spatial statistics





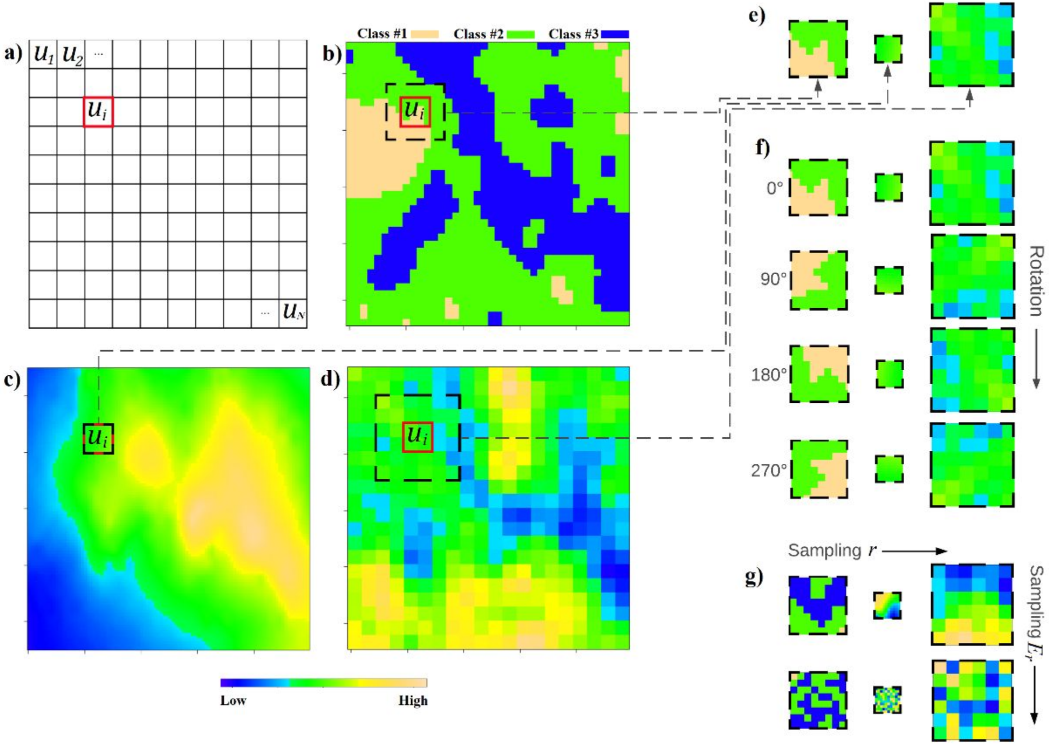


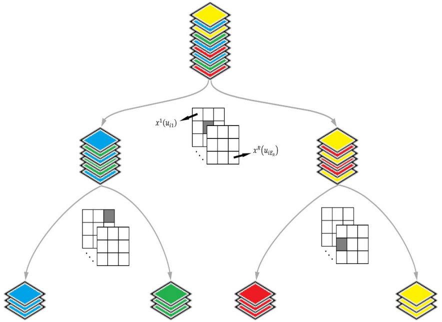


Geostatistical Learning

Spatial Random Forests

- Multiresolution and mixed data
- Supervised and unsupervised learning





$$pat_r(u_i) = [x^r(u_{i1}), \dots, x^r(u_{iE_r})]$$

$$pat(u_i) = [pat_1(u_i), \dots, pat_R(u_i)]$$

$$pat^*(u_i) = [pat_1^*(u_i), \dots, pat_R^*(u_i)]$$

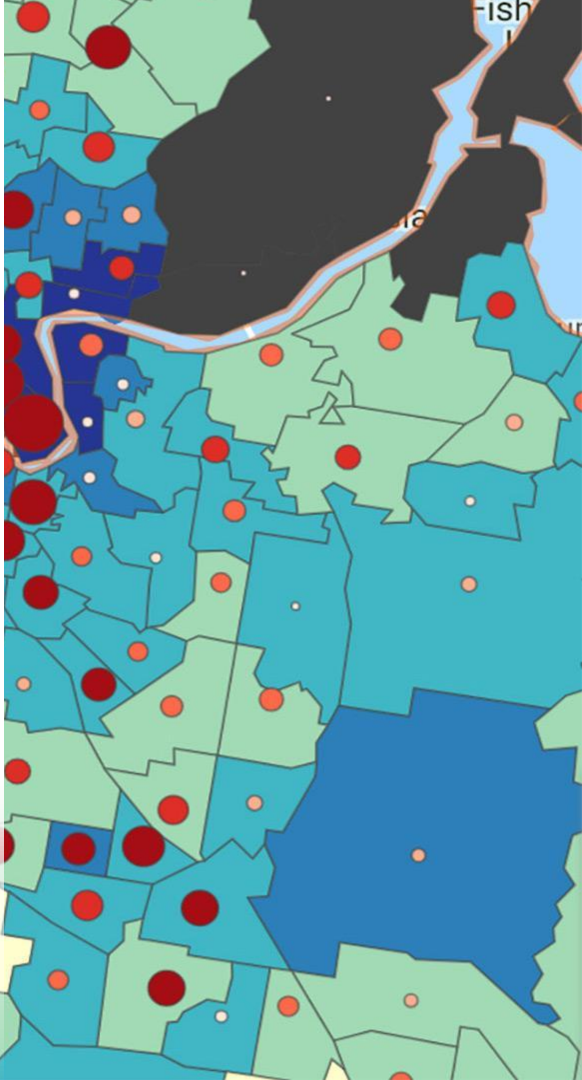
$$\mathcal{D} = \{(pat(u_1), y(u_1)), \dots, (pat(u_N), y(u_N))\}$$

Digital Twins and Research Infrastructures in Australia

The Gemini Principles

Dr Michael Rigby
Data Science Support Team Lead

AURIN
aurin.org.au

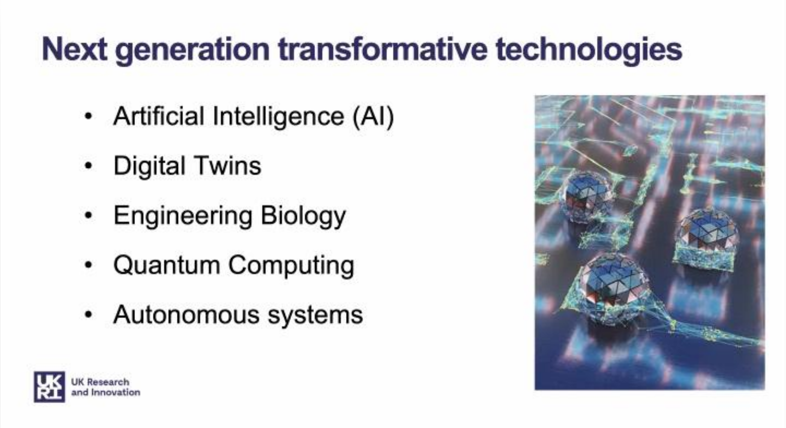


Digital twins are digital representation of cities/towns that are connected to the physical world

Impacts are predicted across various social, economic and environmental dimensions

Given this breadth, a framework is required to guide the development of digital twins

Transformative

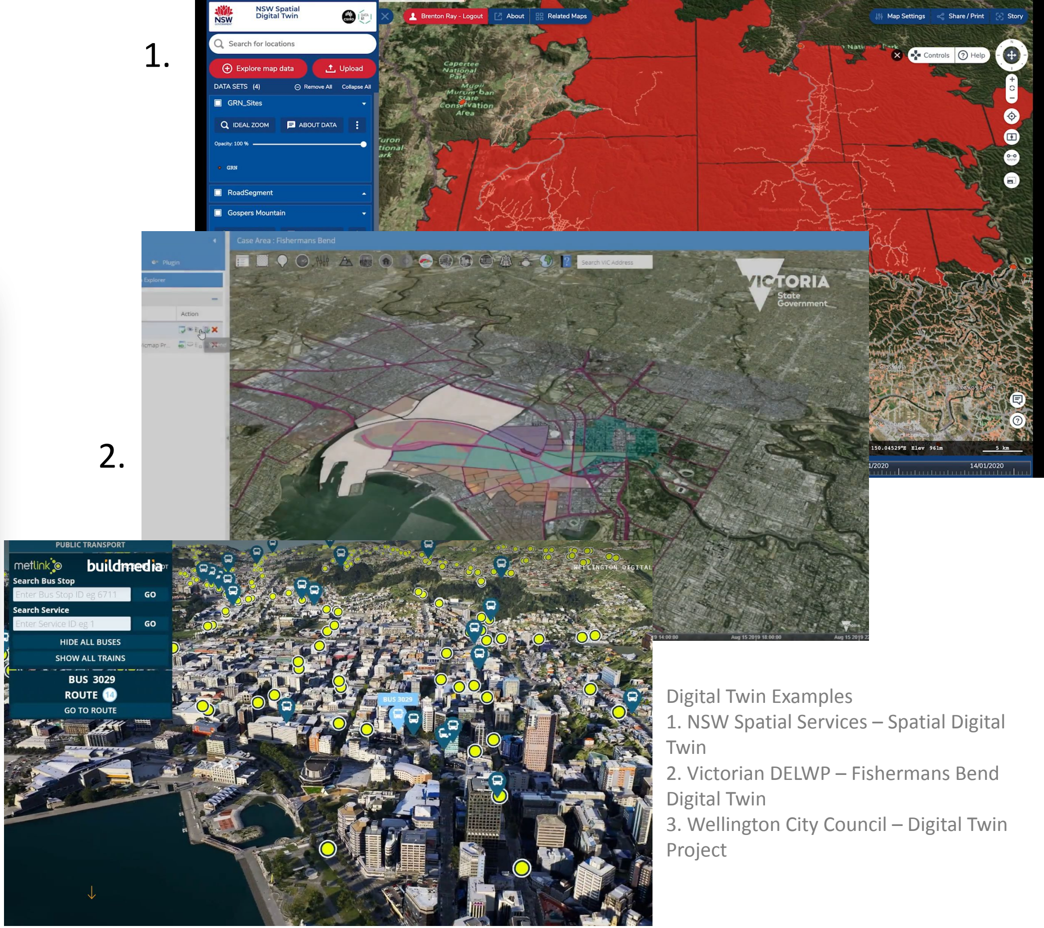


Source: DAFNI 2 Launch Event, July 2021

1.

2.

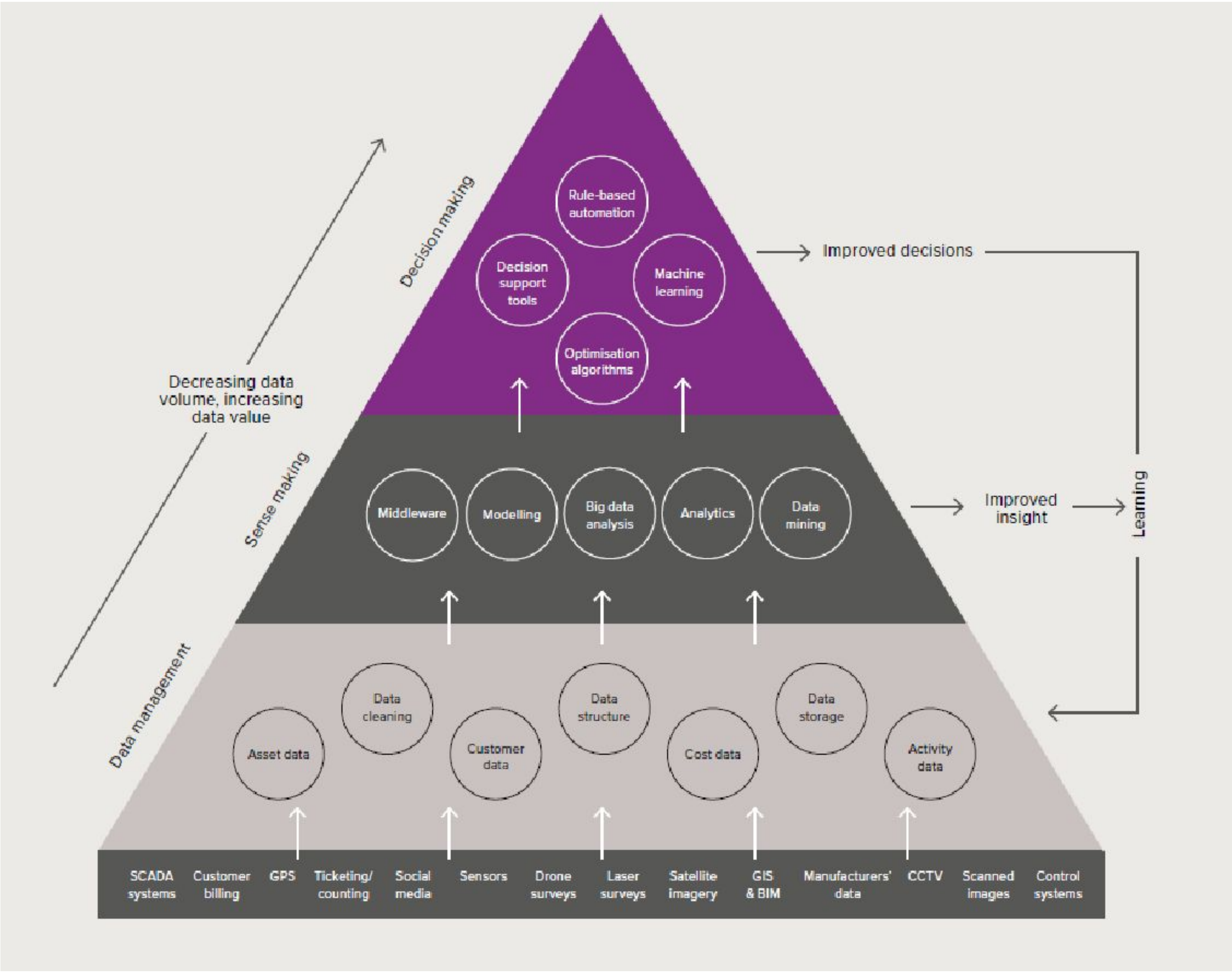
3.



Digital Twin Examples
1. NSW Spatial Services – Spatial Digital Twin
2. Victorian DELWP – Fishermans Bend Digital Twin
3. Wellington City Council – Digital Twin Project

4. FrontierSI
https://frontiersi.com.au/digital-twins/

Information value chain



Source: Bolton, A. et al. (2018)
The Gemini Principles: Guiding values for the national digital twin and information management framework.
DOI: https://doi.org/10.17863/CAM.32260

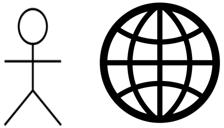
Design process

1. Purpose

Real-time decision making
Dynamic model

Strategic Planning
Static model

2. Requirements



Representation must meet requirements that address the intended purpose

Data – quality (accuracy, scale, granularity, etc.)
Model – quality (parsimony → appropriateness of assumptions, algorithms, etc.)
Visualisation – quality (effective visual codings, etc.)

Human-centred:
ISO 9241-210:2019

Adapted from Bolton, A. et al. (2018)

The Gemini Principles

Guide the development of the framework and digital twin across three key areas:

Purpose	Public Good	Value Creation	Insight
Trust	Security	Openness	Quality
Function	Federation	Curation	Evolution

Source: Bolton, A. et al. (2018) The Gemini Principles: Guiding values for the national digital twin and information management framework. DOI: https://doi.org/10.17863/CAM.32260

Quality

Relates to the appropriateness of the data in relation to requirements

Must be transparent, defined and measured

A minimum standard of quality will be needed, e.g. functionality, security

Success of the digital twin judged of quality of decisions to enables

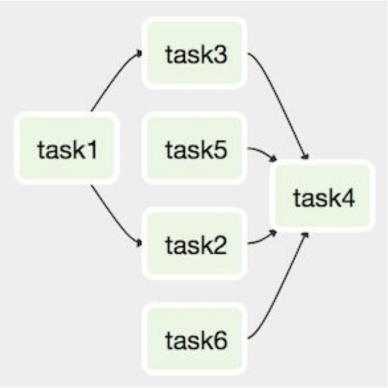
Source: Bolton, A. et al. (2018) The Gemini Principles: Guiding values for the national digital twin and information management framework. DOI: https://doi.org/10.17863/CAM.32260

Data integration

Data management to enable the integration of data

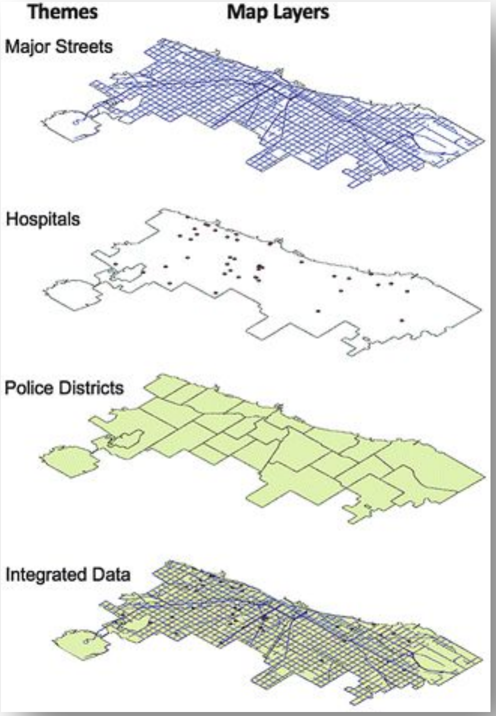
Methods vary dependent on dynamic or static model

Quality needs to be mapped throughout



Harenslak (2019) godatadriven.com

Simplistic view



Balasubramani and Cruz (2018)

Breakout Topics

1. What are the key challenges with big geospatial data? (Sanjeev Srivastava)

1. What are the key challenges in the use of AI in Geospatial Data? (Jens Klump)

1. What are the key challenges to consider when incorporating built/natural environment and social data within Digital Twins? (Michael Rigby)

For use during breakout discussion: https://miro.com/app/board/o9J_l8Yp5Dc=/

Report back: **Breakout 1 -** What are the key challenges with big geospatial data?

1. **Heterogeneity, size & scale, errors, methods and techniques for processing.**
2. **Completeness of coverage, biases in the data (how do you identify them?)**
3. **….**
4. **….**
5. **…..**

Report back: **Breakout 2** What are the key challenges in the use of AI in Geospatial Data?

1. **Need for developing proper spatial machine learning methods**
2. **How to extrapolate from machine learning? Needed in geological exploration**
3. **Need for explainable ML to help understand processes**
4. **Are there suitable methods for dimensionality reduction? (Min/Max Autocorrelation Factors)**
5. **Can we extrapolate to higher dimensions? e.g. from 1D to 2D/3D**
6. **Take into account physical laws (sanity checks)**
7. **Tracking features across scenes/frames**

Report back: **Breakout 3** What are the key challenges of incorporating built/natural environment and social data within Digital Twins?

1. Experience using geospatial data and performing data integration

   Similarities between geology vs urban research and associated models

1. Challenges integrating and establishing built environment data

   Terrain, roads, buildings

1. Social data challenges
   a. Spatial/temporal (patchiness and handling over time)
   b. Connecting these to the built environment
   c. Data availability
   d. Secure/sensitive data ….

Next steps

ABOUT

The intent of the Geospatial Capabilities Community of Practice is to bring people together as a community who, in some capacity, work with geospatial data and make it available for broader use. The group is intended as a forum for sharing knowledge and experiences with finding, accessing, transforming, using and making available geospatial data.

Terms of Reference - Geospatial Capabilities Community of Practice

OVERVIEW

Geospatial data creation and use, and related capabilities

https://sites.google.com/ardc.edu.au/geospatialcapcop/join