Save the tiger in the cloud

Using the cloud in an international research project

HAS green academy

SURF

SAVE THE TIGER THE GRASSLANDS THE WATER

Who are we

Paulo van Breugel
Lecturer/researcher
HAS green academy
Ecology, biodiversity, spatial data analysis & remote sensing,

Robert Jan Bood
Advisor
SURF
Software development, DevOps and cloud technology.

Mark Terlien
Lecturer/researcher
HAS green academy
Data analysis, Geo-ICT, remote sensing, geoscience

David Šálek
Data processing Advisor
SURF
Adoption of public cloud services and cloud-native technologies in research environment

Content

Save the tiger, save the grassland, save the water project

- The study area
- Remote sensing in the project
- Remote sensing desktop or cloud, what do we need?

Save the tiger in the cloud

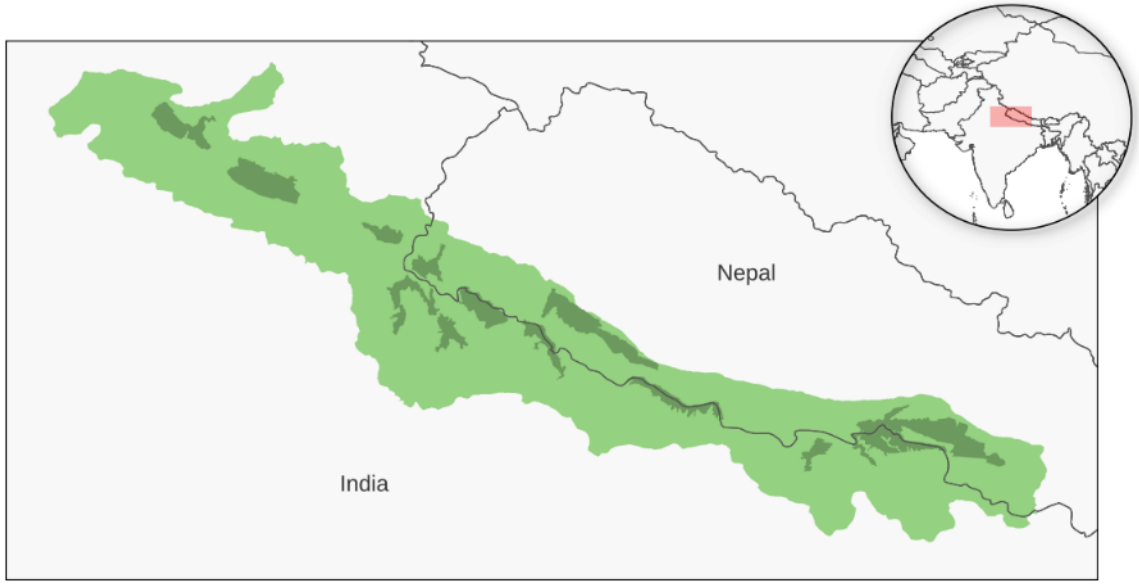
- Project requirements and solutions
- Workflow
- Next steps

Save the tiger, save the grassland, save the water

- A research project funded by the Dutch Research Council (NWO)
- National and international partner organizations (Netherlands, Nepal, India).
- Research institutes, governmental and no-governmental organizations, commercial companies

The Terai Arc Landscape

- The Terai Arc Landscape (TAL) at the foot of the Himalayas are the most important conservation area of tigers.
- There are approximately 900 wild tigers alive (out of 4000).
- Habitats are under pressure due to climate change and anthropogenic activities.



The Terai Arc Landscape

- The Terai Arc Landscape (TAL) at the foot of the Himalayas are the most important conservation area of tigers and other wildlife.
- There are approximately 900 wild tigers alive (out of 4000).
- Habitats are under pressure due to climate change and anthropogenic activities.






The Terai Arc Landscape

- The Terai Arc Landscape (TAL) at the foot of the Himalayas are the most important conservation area of tigers and other wildlife.
- There are approximately 900 wild tigers alive (out of 4000).
- Habitats are under pressure due to climate change and anthropogenic activities.

The Terai Arc Landscape

- The Terai Arc Landscape (TAL) at the foot of the Himalayas are the most important conservation area of tigers and other wildlife.
- There are approximately 900 wild tigers alive (out of 4000).
- Habitats are under pressure due to climate change and anthropogenic activities.



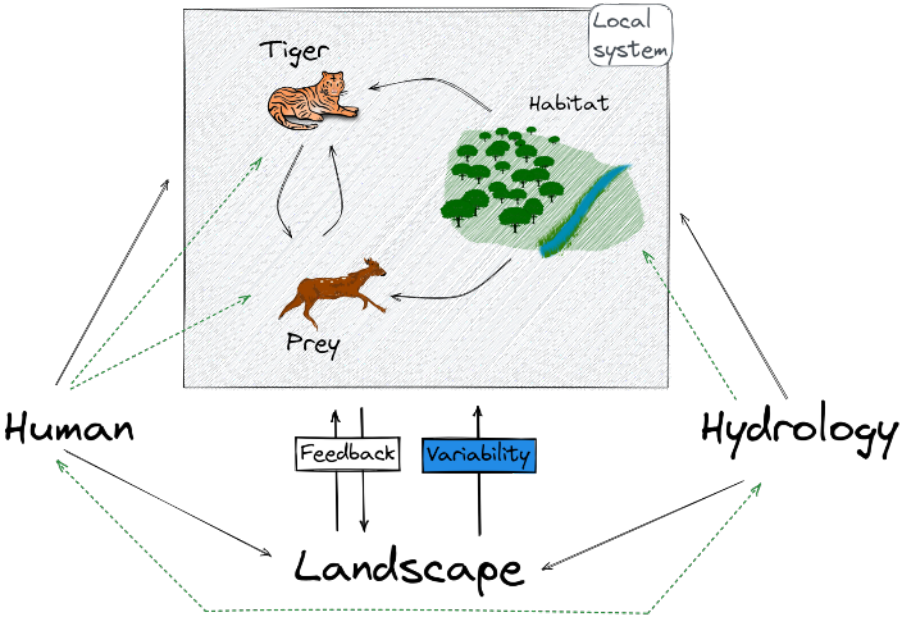



Project goal and objectives

- Insights to support a sustainable management of the habitat of the tiger and its prey in the Terai Arc Landscape (TAL).
- To understand, model and monitor the spatio-temporal dynamics of the habitat in nature reserves in the TAL.

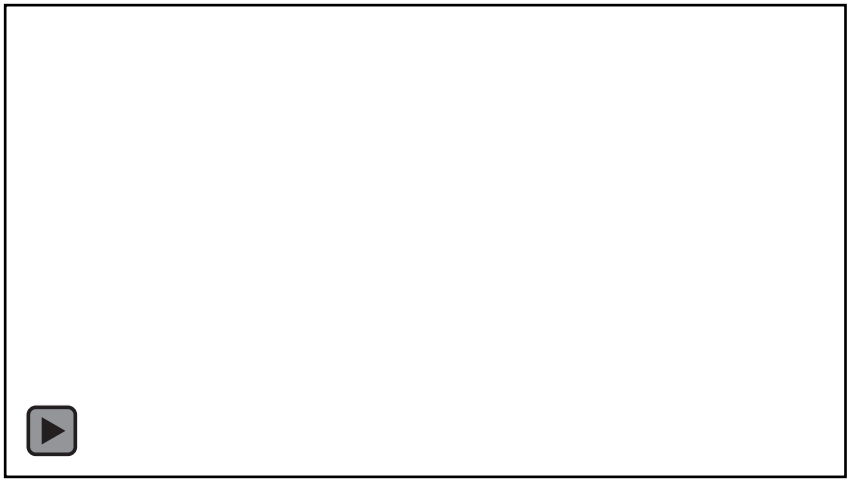
↳ What are the natural dynamics of the vegetation in the TAL?

↳ What are main anthropogenic factors influencing these natural dynamics?



Remote sensing

as a solution

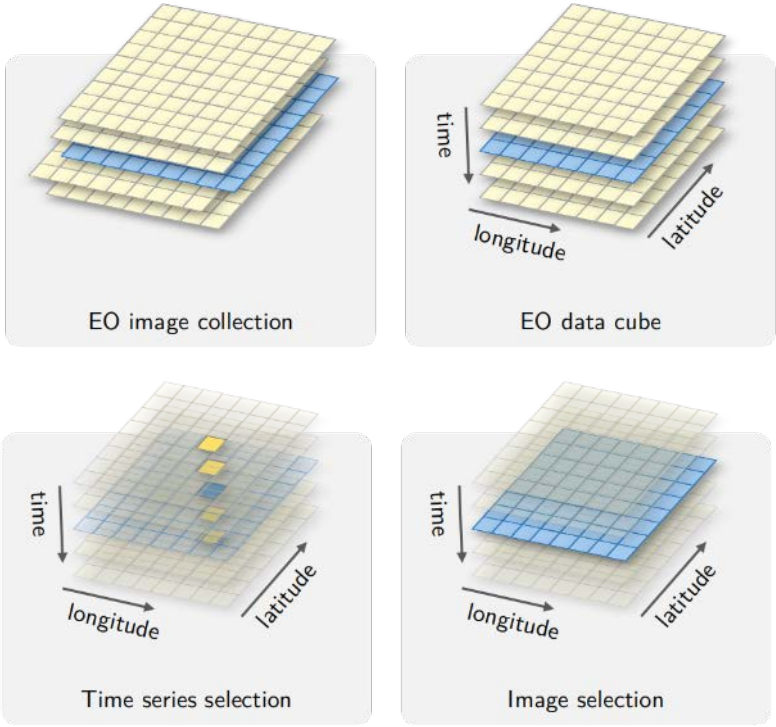


- Cover large areas
- Low to high frequency
- A lot of high-quality open data and commercial products
- Highly consistent data

Remote sensing

Data products

- Multi-dimensional (time, space, bands).
- Can be complex to retrieve the data you need.
- Very large data volumes



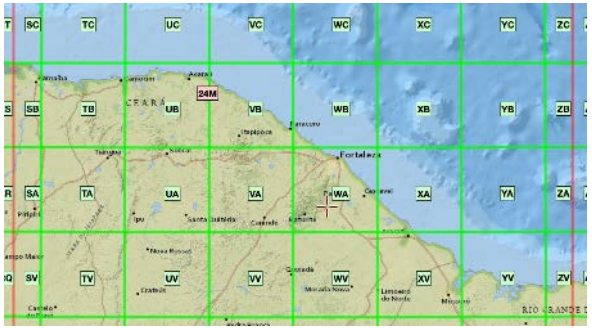
Source: sits

Remote sensing

Data products

- Each product with its own specifications and characteristics (tiling, resolution, frequency, number and with of spectral bands, data format, metadata description)
- Specifications change regularly

WRS-2 tiling system used by Landsat-5/7/8/9 images (Source: INPE and ESRI.)

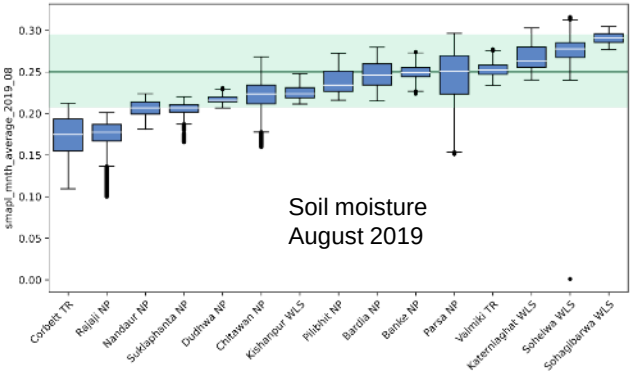


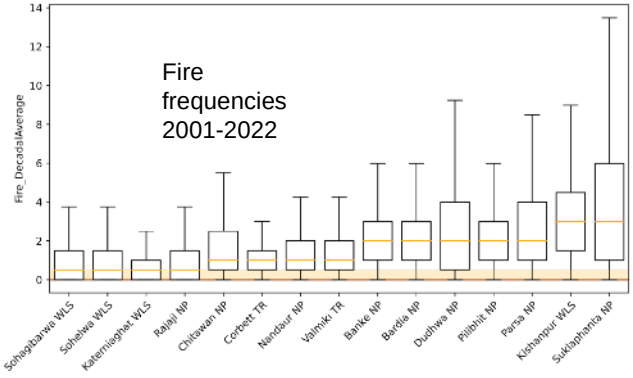
MGRS tiling system used by Sentinel-2 images (Source: GISSurfer 2.0

Remote sensing

environmental patterns

How do environmental conditions differ between and within the national parks?

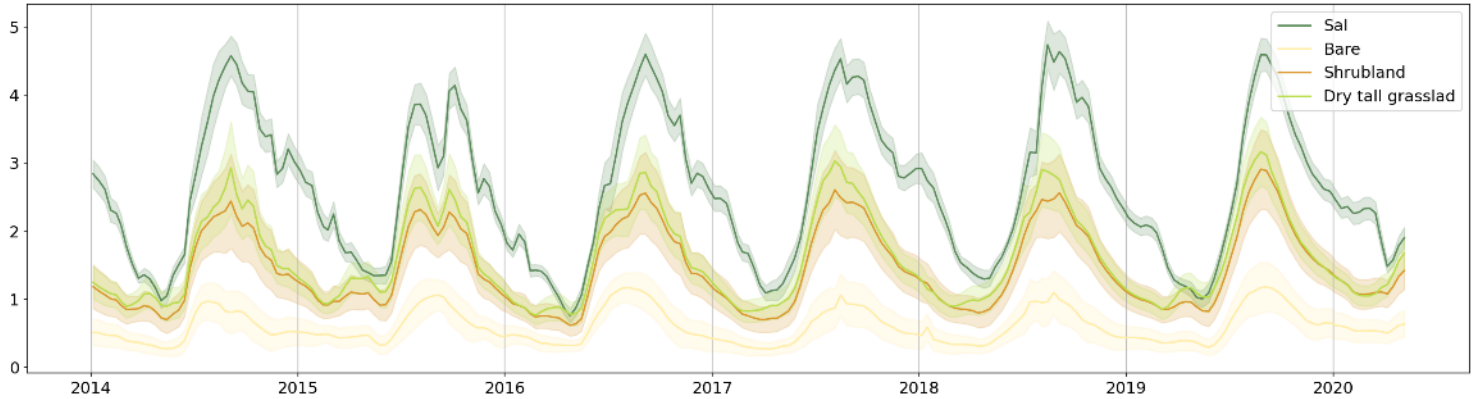




Remote sensing

Time series analysis

Variation of vegetation and hydrological characteristics over time, identify seasonal patterns and detect anomalies



LAI variation between 2014-2020, for four different vegetation types in Bardia National Park.

Remote sensing

Classification vegetation and land use

Fieldwork

+

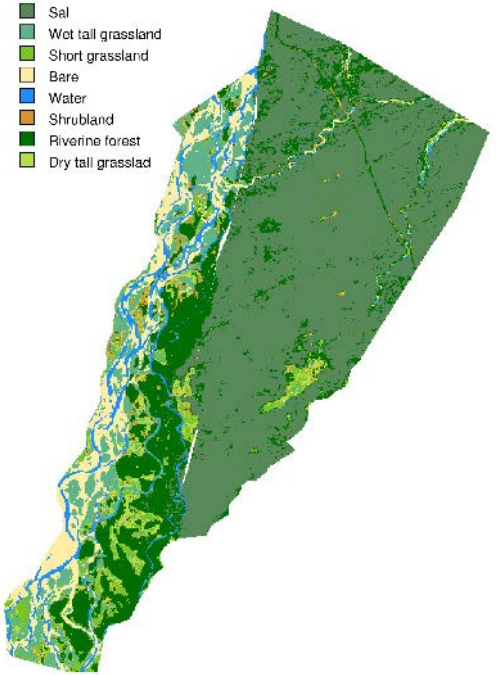
satellite images

+

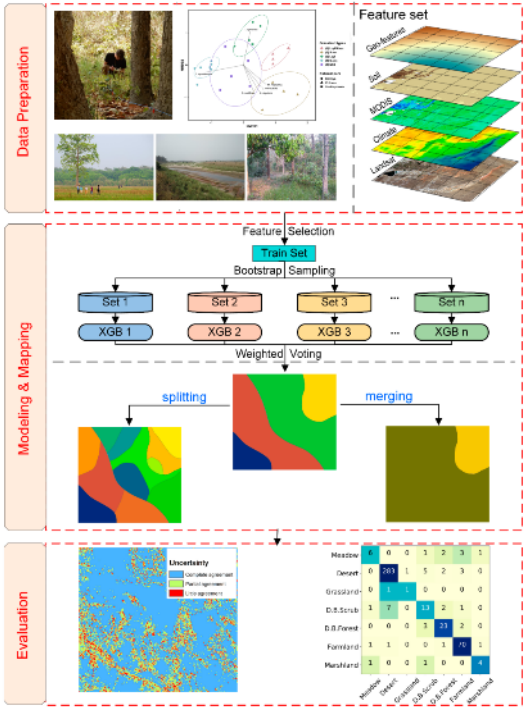
ancillary data

+

Machine learning



Source: Bijlmakers 2023



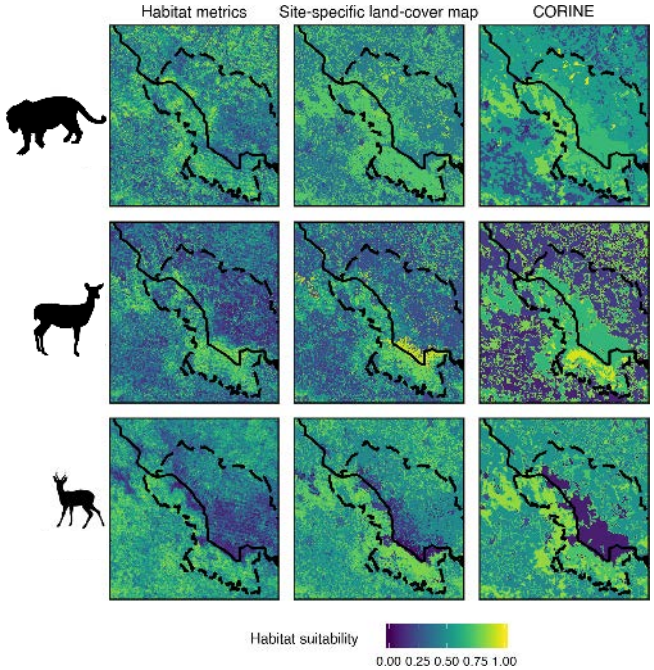
Adapted from: Zhang 2019

Remote sensing

Human activities, ecology and hydrology at the landscape level

- Linking habitat patterns and changes to the distribution of prey species of the tiger (trend analysis)
- Identifying and detecting influences of anthropogenic interventions on vegetation and hydrological dynamics (effect monitoring).

- Standardisation
- Automation



Adapted from: Oeser et al. 2020

Remote sensing on the desktop

What are the challenges?

- Large data volumes, complex data setup
- Working across countries, organizations and disciplines.
- Working with different levels of expertise (scientists, private parties, students).
- Lack of consistency in tools (libraries, software packages), workflow and methods
- Hardware requirements (scaling)

What do we need?

- An easy-to-use data storage platform for our partners and students, with analysis-ready data (ARD) for cloud computing and use on desktop.
- Granular access to data and resources on per-user basis.
- Easy to integrate existing analytical tools and solutions (e.g., Open Data Cube, Python and R libraries) and mix different data sources (e.g., ARD collections on AWS, MS Planetary computer, Planet Planetary variables)
- Gain experience and insights in using the cloud for research and student projects.

What do we need?

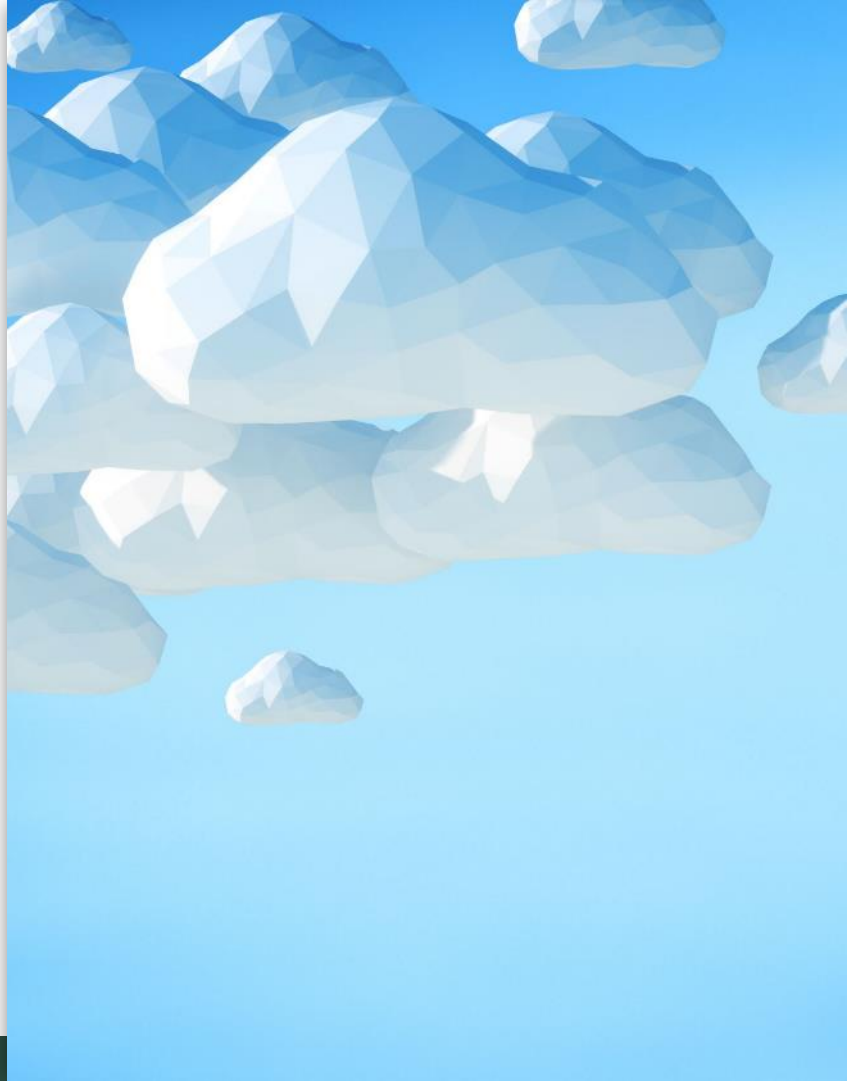
	Existing platforms
Easy storage and retrieval of data.	⚠
Enable partners to access the data from anywhere.	✓
Full control over data, granular access.	✗
Consistant preprocessing steps, well documented data provenance and management of versions of data products.	⚠
Low maintenance, fits within existing stack of products.	⚠
Easy to develop, share and document tools, routines, workflows	⚠
Cater to different types of users (beginners, experts)	✗

Save the tiger on Azure

- Where do we stand
  - Data storage solution for Sentinel-2
  - Scripts and documentation/tutorials being developed
  - Student project to use and test data storage and python analysis framework
- Next steps
  - Making it land in the organization and build up expertise
  - Getting partners in the project onboard (training)
  - Expand with other RS products

Cloud research consultancy

- Customisation for complex research project
  - https://www.surf.nl/en/services/cloud-research-consultancy
- Co-development project HAS and SURF
  - Proof of concept solution
  - Azure Cloud services



Project requirements

Scalable and Elastic

Data size varies

Changing compute requirements

Experiment with different ML models (training/inference)

Managed/serverless services

Low maintenance overhead

- No OS updates & less configuration
- Focus on functionality

Easy integration between components

- Faster setup

Preconfigured analysis environment

Different levels of expertise

Consistent environment for students

Packages and libraries installed

Importance of elasticity

- 1 Sentinel 2 image (RGB bands only) is 375 MB
- Complete Sentinel image
- TAL-region is 50.000 km2 & 700km long (20-60 images)
- Seasonal changes: analyse each season, month, week
- Temporal Changes (1-10 years)
- Cloud cover: patch multiple images
- Experiment with deep learning

Solution overview

Frontend

- GeoServer
  - Share, explore, and edit geospatial data
- Jupyter notebook environment
  - Analyse data
  - Machine learning

ML Studio

App Service

Backend

- Satellite Image API
  - Query Sentinel API
  - Start download requests

Functions

Storage Queue

Data Storage

- S3 Object store
  - Raw image files
- Metadata Database
  - For example: timestamp, cloud coverage, vegetation, snow, etc.

Blob Storage

Database PostgreSQL

Example workflow

GeoServer

Explore available data

Satellite image API

Request images

Download tasks queued

DB & object store (automated)

Unzip & convert image data

Retrieve metadata

Notebook environment

Analyse data

ML training

Inference

Serverless compute

On-Demand compute

Next steps

Students projects

- Download more data
- ML workloads
- Feedback

Review solution

- Usage & costs
- Maintainability
- Pilot to production