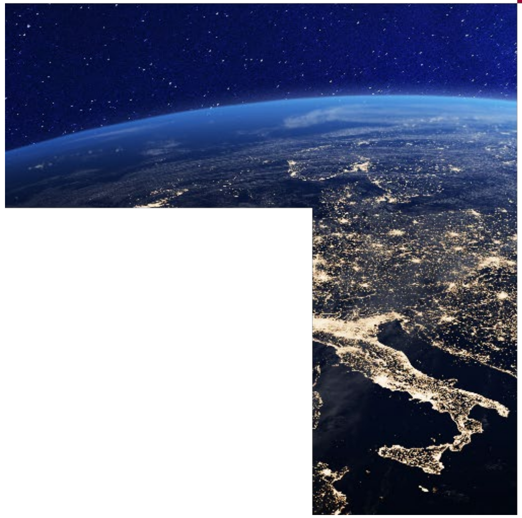
The lessons learnt and roadmap to Digital Twin Earth: Food Systems precursor project

Dr Chandra Taposeea-Fisher

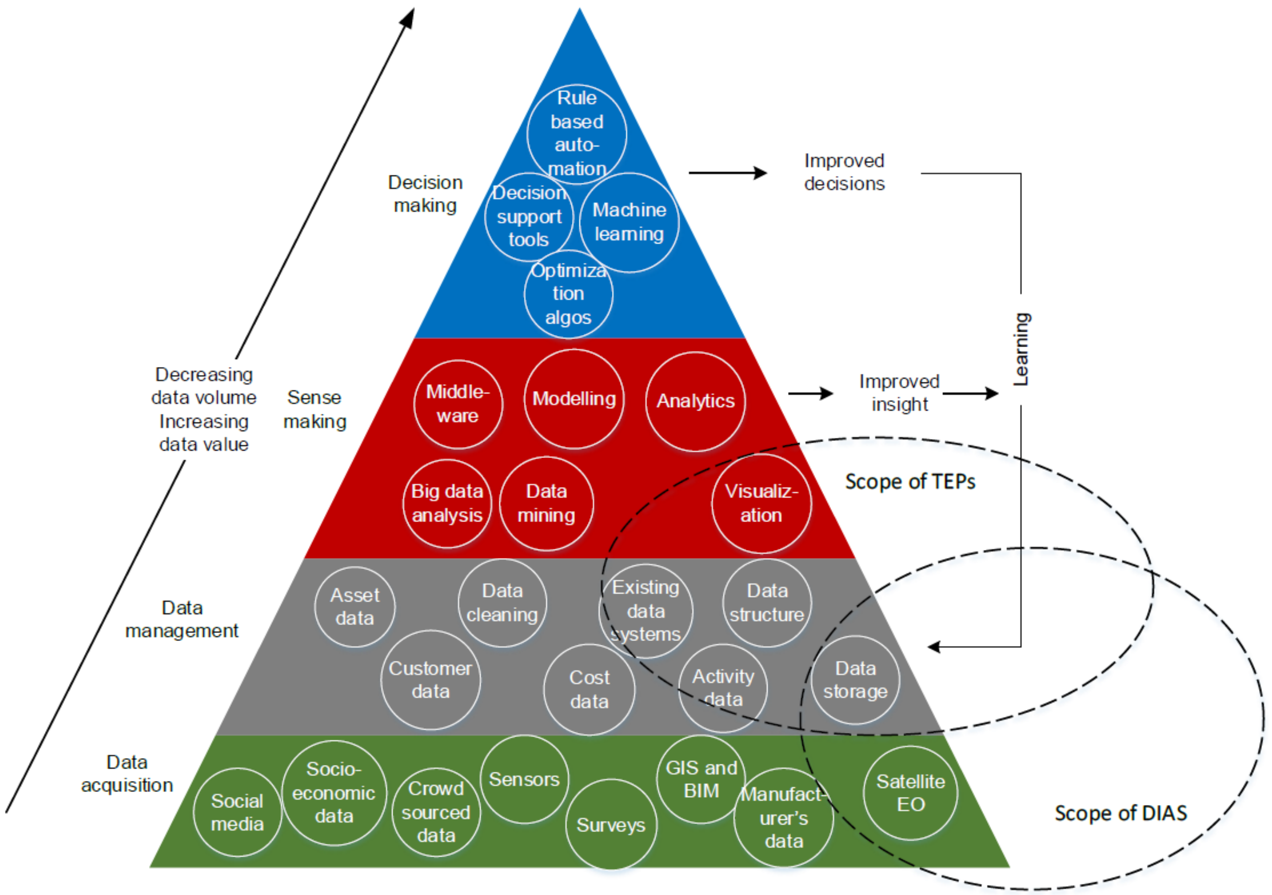
Senior Consultant | Earth Observation



living planet symposium | BONN 2022 | #LPS22

About the project

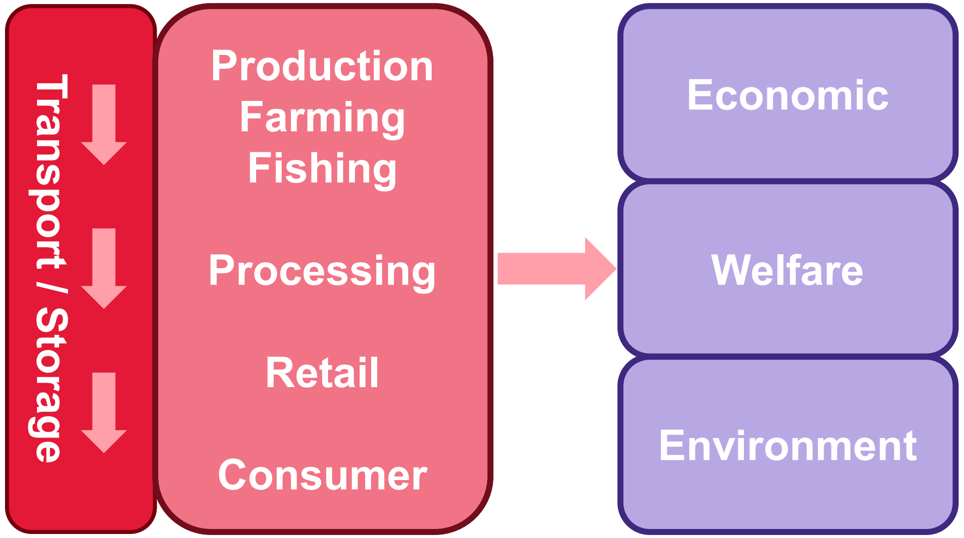
- Focusing on supporting ESA in the definition of the concept of a digital twin
  - **Horizontal component**: focuses on common elements of DTE architecture and shared building blocks of end 2 end implementation
  - **Vertical component**: focuses on specific applications, science domains and use cases, to complement horizontal components.
- Worked towards a roadmap for recommendations to address current scientific / technical limitations
- This precursor focuses on a Food System Digital Twin as a use case



© 2022 CGI Inc.

Our case study: Food Systems

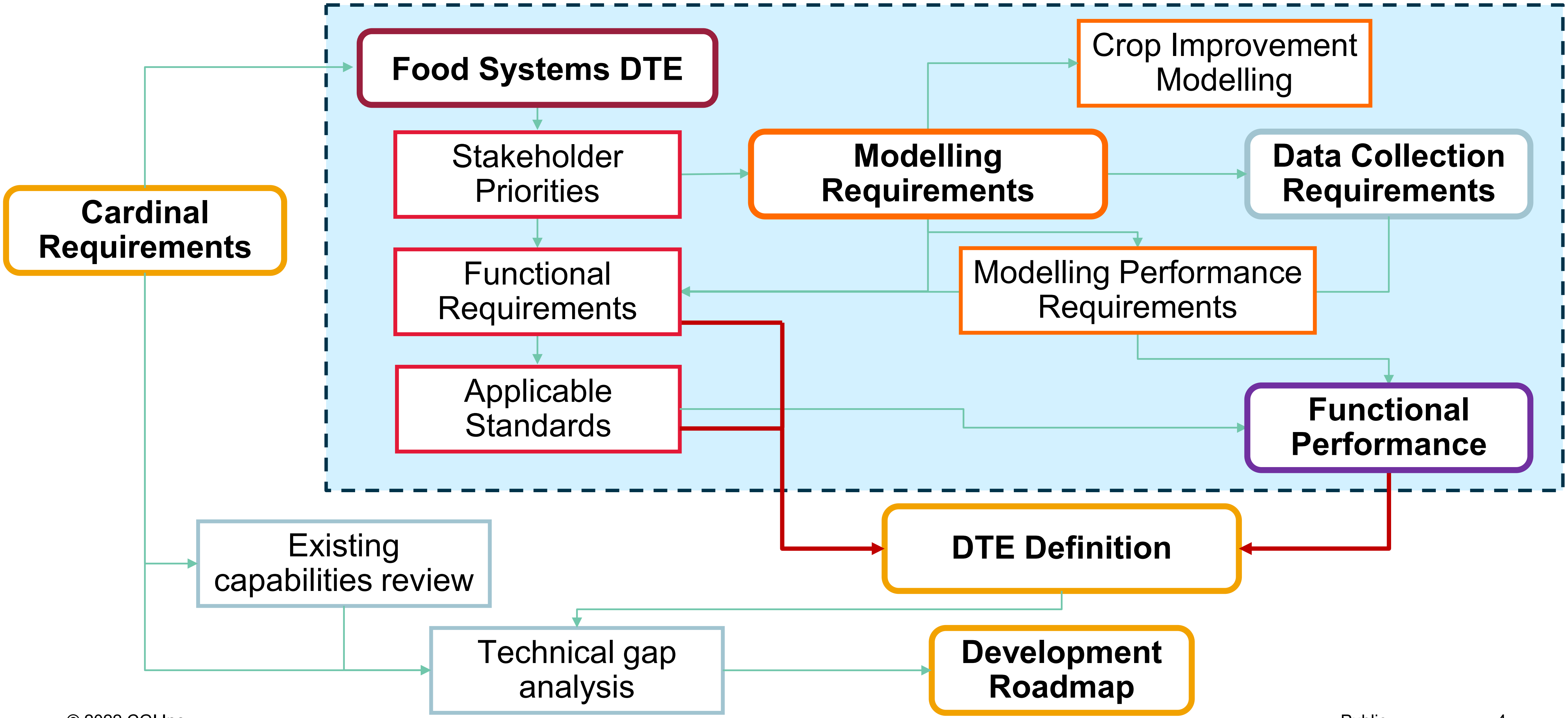
- The economic importance of food systems is considerable.
- Extending well beyond farming, DTE provides an opportunity to further develop and enhance this information:
  - Predicting extremes more effectively
  - The ability to work at a range of scales
  - Integration of crop modelling at a range of scales
  - Better integration with related modelling domains & socio-economic data



Our use case: Looking at how extreme precipitation events (from Oxford University) interacts with global crop models, both physical and socioeconomically (from IIASA) and building a sustainable infrastructure behind it (from CGI)

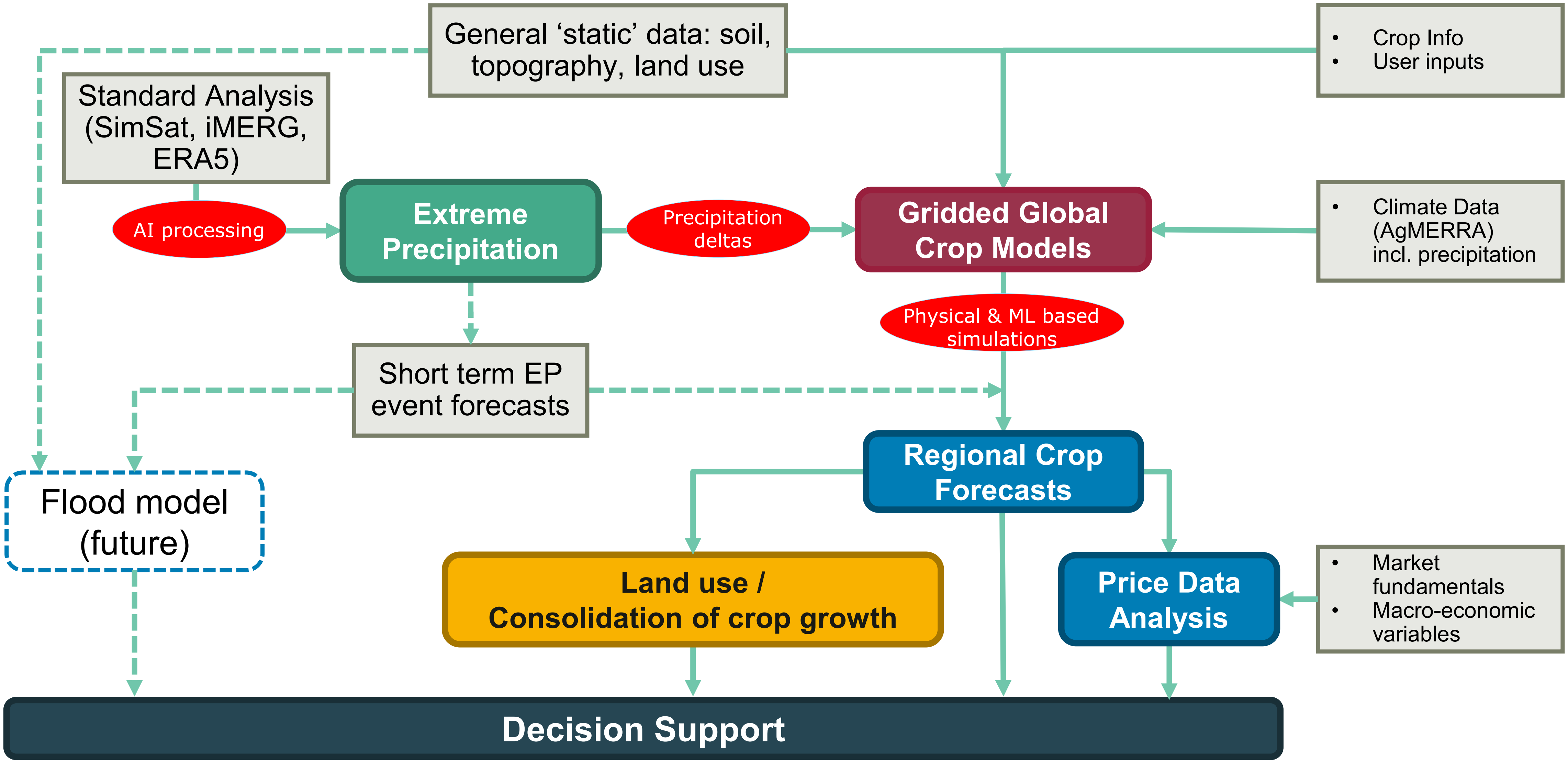
© 2022 CGI Inc.

Theme Definition & User Requirements



© 2022 CGI Inc.

Model Linkages



Stakeholder Inputs (workshops in January & June 2021)

Food and Agriculture Organization of the United Nations
National Centre for Earth Observation NATURAL ENVIRONMENT RESEARCH COUNCIL
UKRI Biotechnology and Biological Sciences Research Council
Met Office
UKRI Science and Technology Facilities Council
KING'S College LONDON
Helmholtz-Zentrum Geesthacht Centre for Materials and Coastal Research
UKRI Natural Environment Research Council
arqiva
CABI
NIAB

Their interests

High Level / Food Systems Specific:

- Support end-to-end analysis of food supply chain vulnerabilities
- Support uptake of new policy developments

Modelling Requirements:

- Allow creation of scenarios to compare impacts of different parameters/data-sets
- Need for detection of sub-grid extremes locally initiated, but also the need for crop modelling on plot scale and larger scale

Data/Functional Requirements:

- Provide access to reanalysis historical data
- Provide links between yield data outputs and commodity price data
- Visualise relevant data in a time series and at a range of scales

© 2022 CGI Inc.

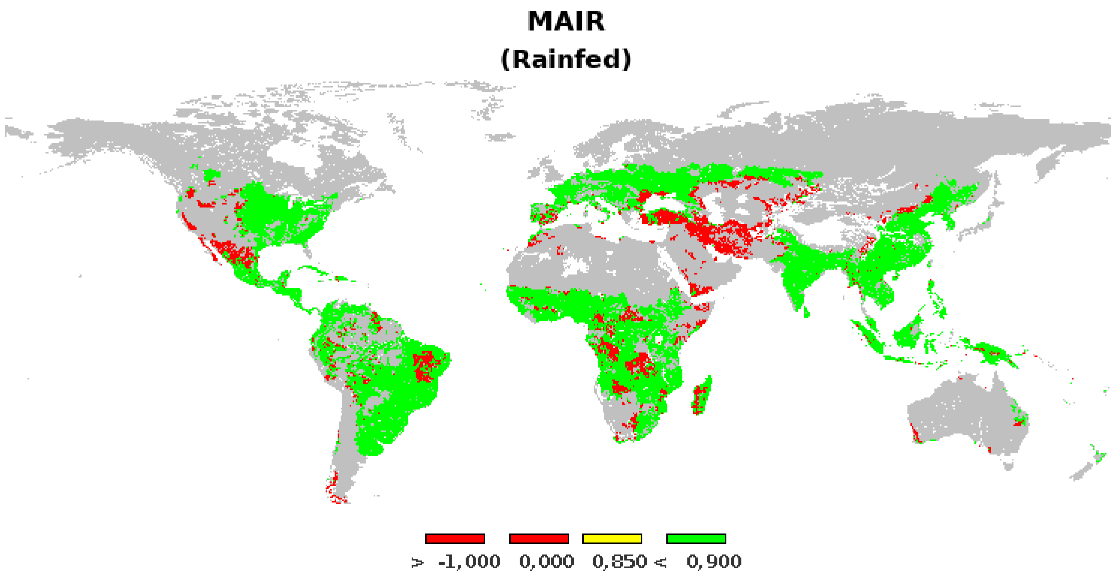
Data and Models used
Global Gridded Crop Models (IIASA)

EPIC-IIASA (Environmental Policy Integrated Climate Model)

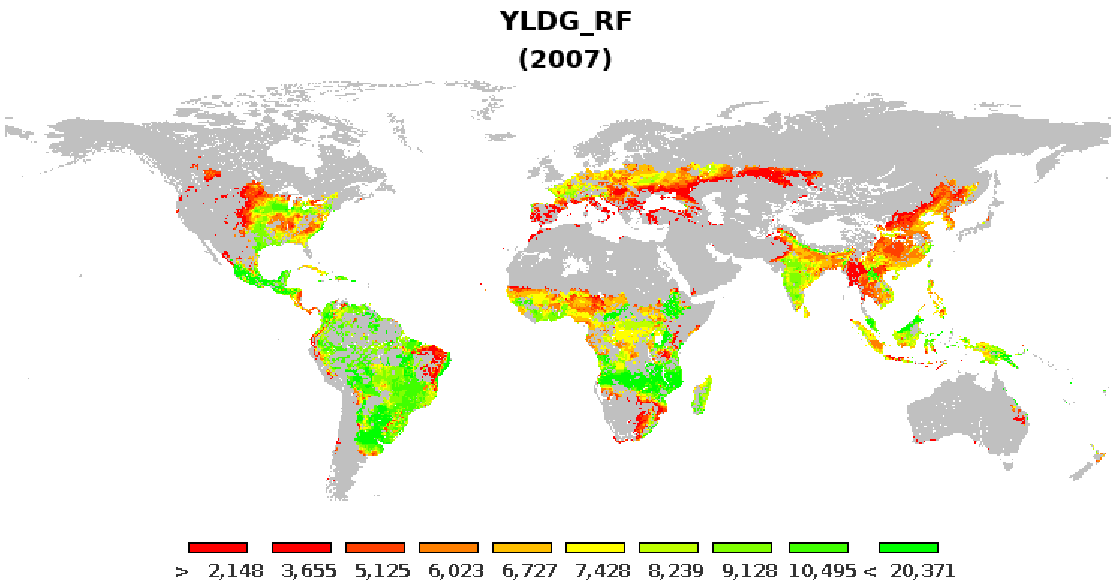
- Simulates agricultural activities and interactions within ecosystems daily
- Run separately for each simulation unit, crop and water management system (rain-fed or with sufficient irrigation)

AgMERRA dataset (driver for EPIC)

- Modern-Era Retrospective Analysis for Research and Applications (MERRA) – Agriculture version
- Cover 1980-2010 period



Maize production allocation (rainfed) providing for maximum land sparing under constraints of this study. Green color denotes the used land, and red color depicts the land that is saved and given back to nature.

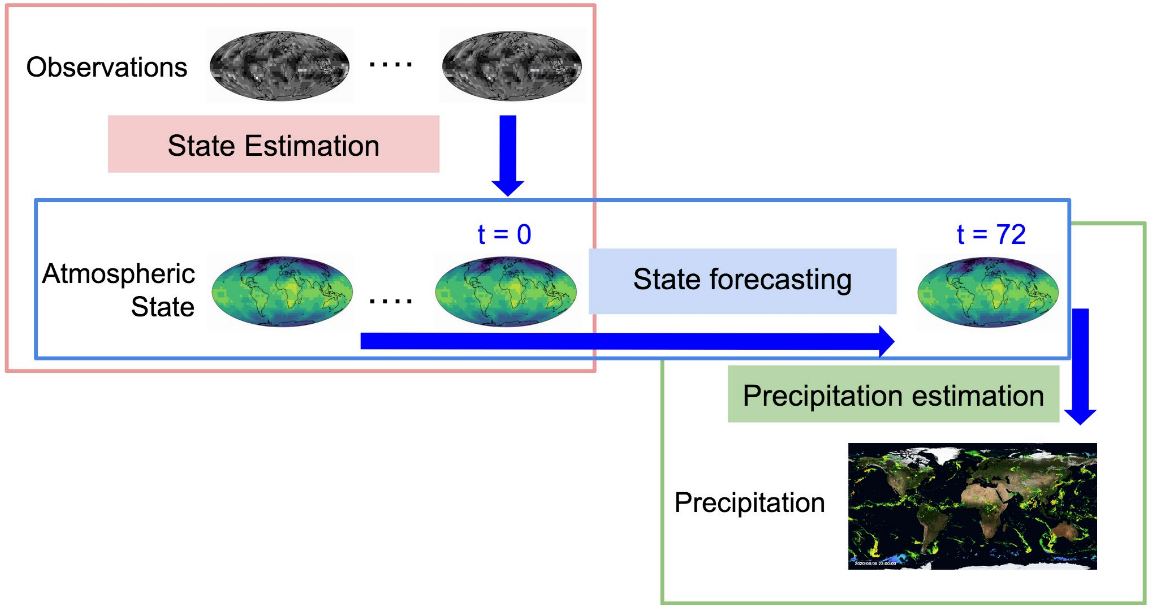


Maize yield potential (rainfed) as estimated by the EPIC model for the year 2007.

© 2022 CGI Inc.

Data and Models used
Precipitation Models (Oxford)

- Construction of the AI derived precipitation product to be used from:
  1. European Centre for Medium-Range Weather Forecasts (ECMWF) simulated satellite data (SimSat);
  2. ECMWF Re-Analysis, 5th edition (ERA5) reanalysis product; and
  3. IMERG global precipitation estimates.
- Three separate ML models used to predict future precipitation from a given SimSat image, referred to as:
  - *State estimation;*
  - *State forecasting;*
  - *Precipitation estimation.*
- To use this pipeline in informing extreme precipitation effects on crop production:
  - There is an adaptation of the ERA5 and IMERG precipitation products
  - Needs to match the AgMERRA meteorological fields expected by the EPIC model.



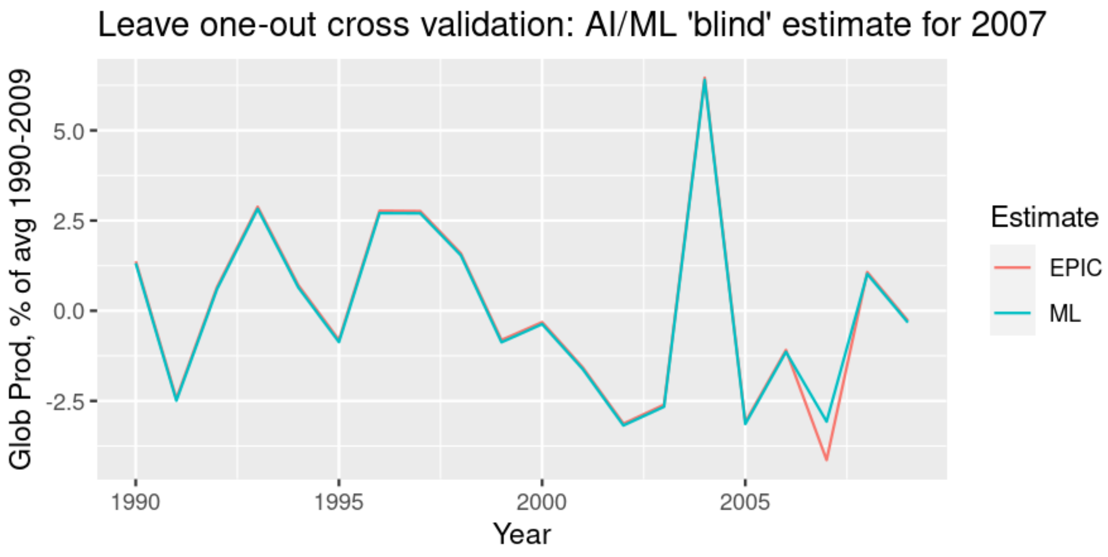
Precipitation prediction system overview. The three steps -- state estimation, state forecasting and precipitation estimation -- are highlighted by different colours.

© 2022 CGI Inc.

Model Integration

Investigate the use of AI/ML for the estimating the adverse effects of extreme weather at global / regional scale in the first instance

- Compared extreme gradient boosting ML with equivalent physics based EPIC model
- Focus on 2007, the year most representative of extremes in global maize growth
- ML trained on historical weather 1990-2009 (excluding 2007) - comparison for global maize production.
- Modelled 2007 production decrease was less pronounced with ML (blue) compared to biophysical EPIC (red)
- Using ML for weather extremes on crop yields more difficult than original ML application of downscaling known yields under known weather to a finer spatial resolution



© 2022 CGI Inc.

Model Integration

Extend the use of AI/ML techniques to investigate impact at a local scale, in the context of an extreme weather event.

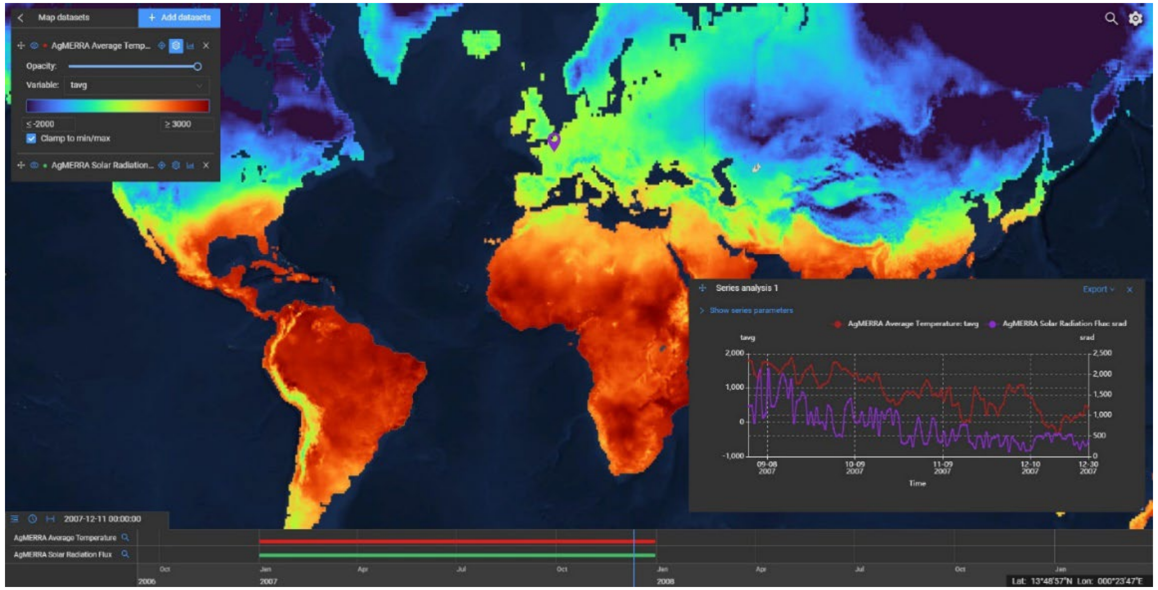
- Running issue of a mismatch of models available
  - Oxford modelling was focused on predicting discrete classifications of precipitation intensity in order to alleviate the large class imbalances found in global precipitation probability distributions
  - IIASA requires a much more aggregated and discrete input (wet days/cell, annual precipitation)
  - Alternative methods investigated (i.e., ecPoint) – ran risk of that it would smooth out the temporally localised, probabilistic extremes
  - Ultimately little purpose to use local precipitation maxima unless the client model has the resolution to make appropriate use of the new information.
- A further research and operational challenge lies in bridging the appropriate timescales between precipitation forecasting (days) and crop modelling (months).
  - Future work could explore the use of such models for improved crop modelling, particular in the case of blocking events which can have severe regional weather implications (such as drought) and affect whole growing regions

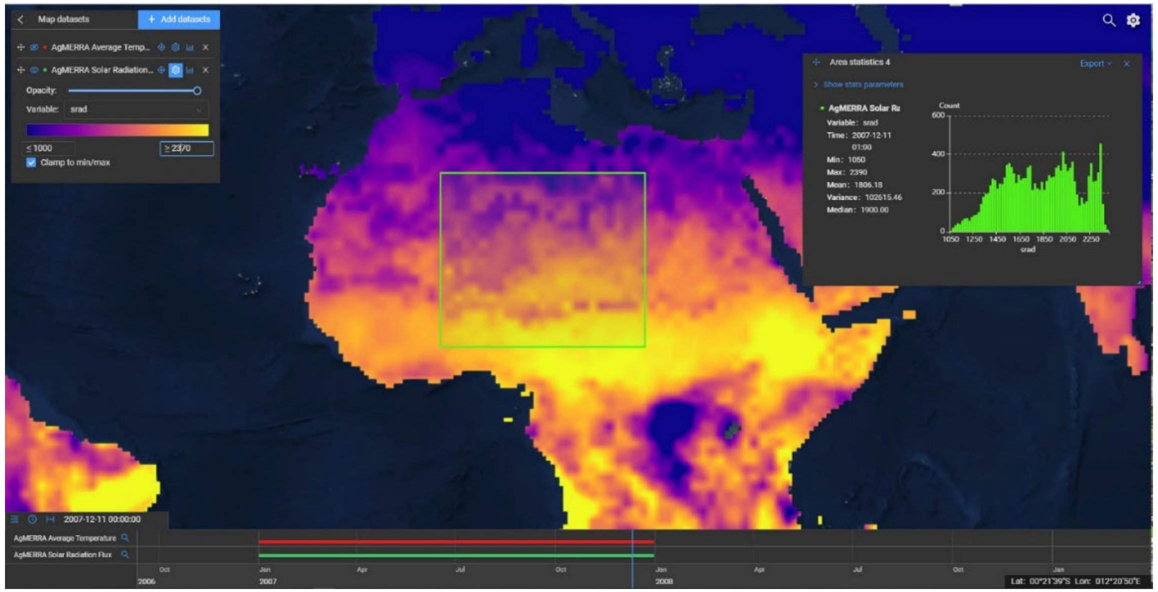
© 2022 CGI Inc.

Visualisation

Visualization and analysis of the climate:

- Discover the available datasets
- Add datasets and modify the layers display order, visibility and opacity
- Extract dataset statistics over an area for a specific instant in time





Infrastructure of Precursor

Preliminary architecture for Food Systems DTE:

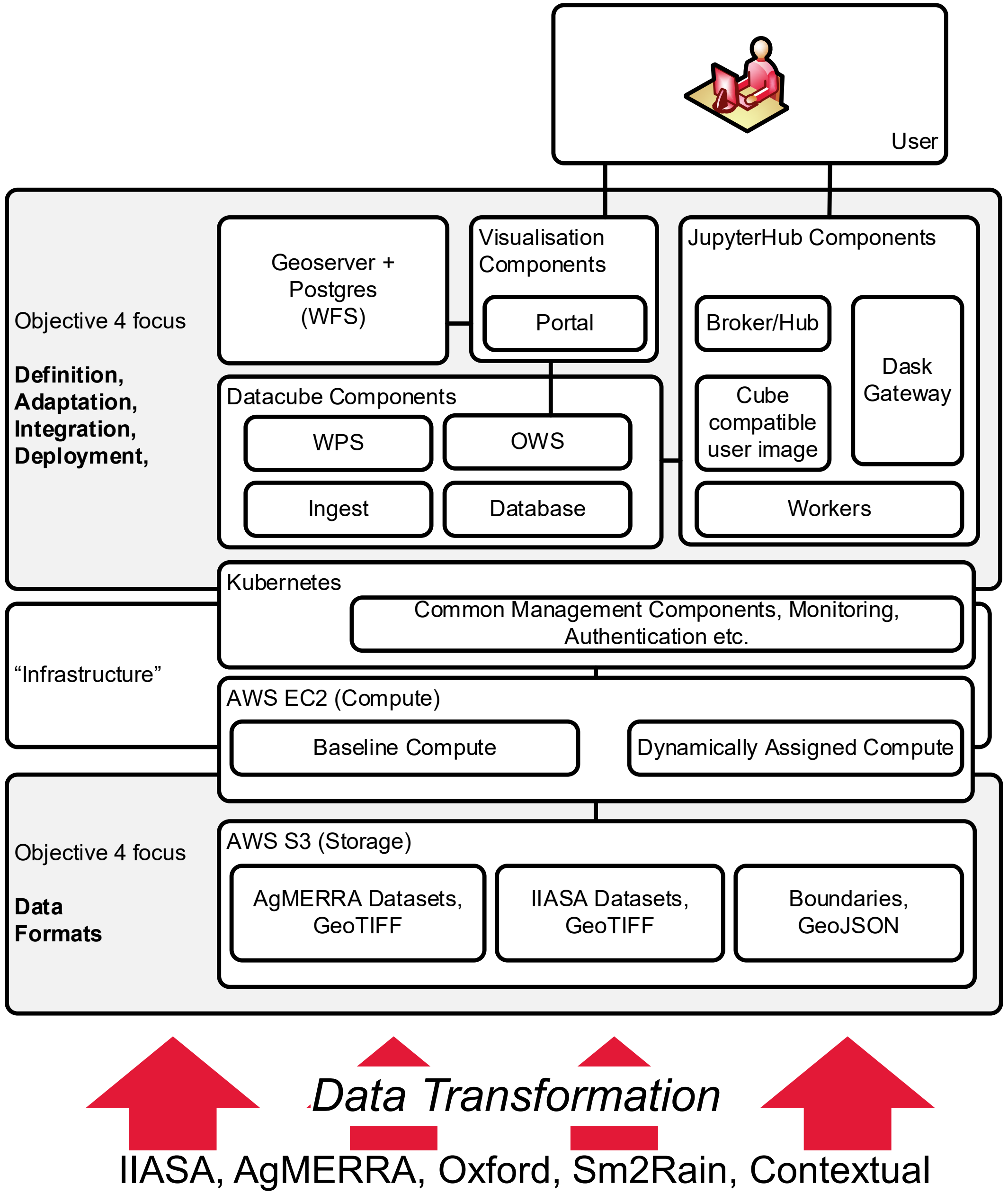
- Investigating data - present data that is amenable to the visualisation
- Deal with the differences between different types of data
- Future proofing it for an operational digital twin

This was done through:

- Bespoke data transformation
- Datacube, on top of Kubernetes, enabling WxS
- Visualisation Application
- Support for jupyter notebooks/data analysis tooling across cluster

Challenges are primarily:

- lack of coherent documentation (FOSS not always maintained)
- and then choices about the data access / transformation for given processing



© 2022 CGI Inc.

Key lessons learnt from demonstration phase

- Simplifying model interfaces
  - Through standardized datasets
  - Use of AI & machine learning to simplify things
  - Possibility of models being shared and compared by different users

- By standardising data, able sharing common concepts between domains.
  - Go beyond food systems
  - Interact with other digital twins

- A significant challenge, particularly for extremes, is the identification of training data; and the assertion and tracing of training data through the system.

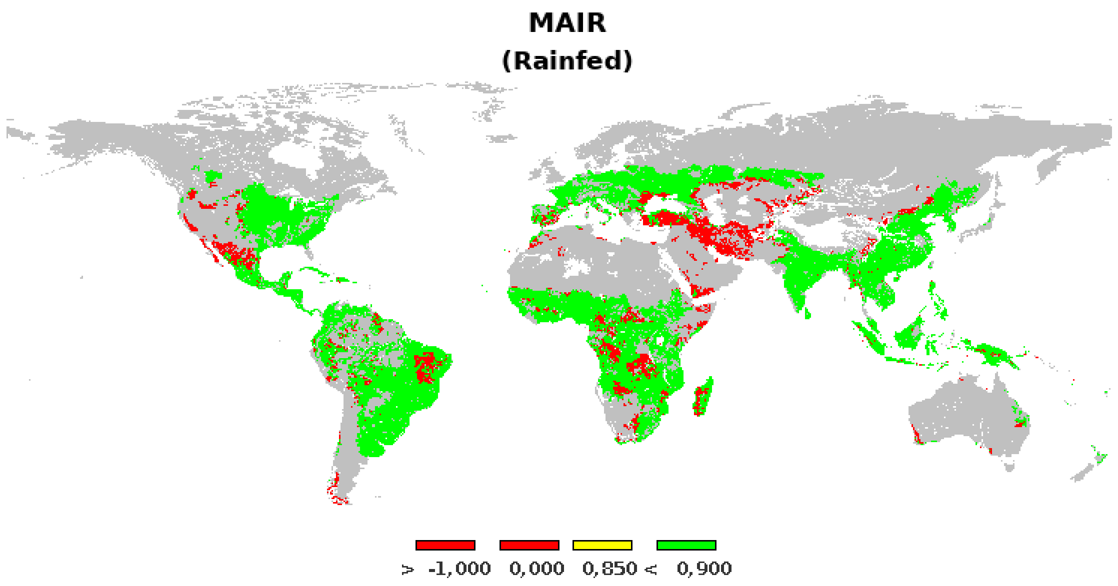
© 2022 CGI Inc.

Policy implications

- Impact of (short term) extreme weather on global maize production under (long term) optimal land allocation
- How would the optimal land allocation need to change in order to be resilient to such weather shocks?
    - Initial estimation of saved land (20.5%) cannot be achieved for the extreme year 2007 whilst maintaining global maize production level
    - On current estimates, only 13.8% of cropland can be saved

Further work:

- How would one go about implementing this?
- How do address traceability? AI/ML often a black box
- How we ensure data is litigation worthy for policy making decisions?
- How we federate data? What happens with proprietary data?

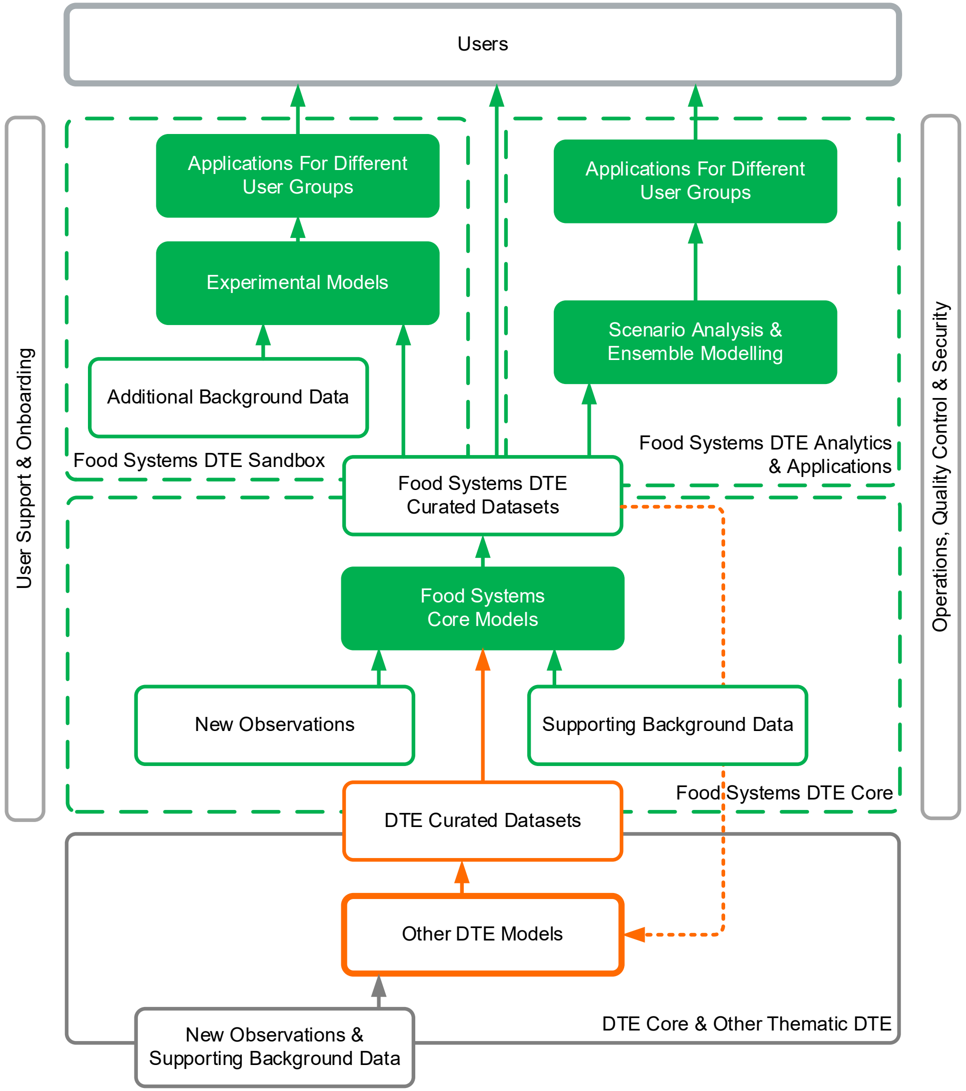


Maize production allocation (rainfed) providing for maximum land sparing under constraints of this study. Green color denotes the used land, and red color depicts the land that is saved and given back to nature.

© 2022 CGI Inc.

DTE Architecture Definition & Functional Requirements

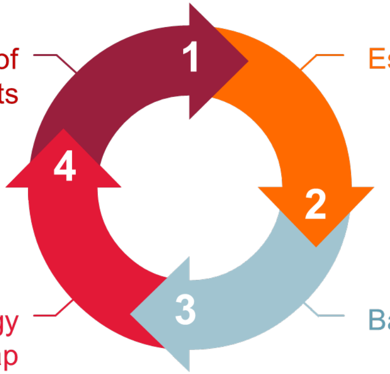
- Users: Research / Commerce
- **Sandbox & Own Algorithms & Data**
- Drawing on Curated Datasets but not only curated datasets



- Users: Policy, Core
- **Scenario Analysis & Reporting**
- Controlled, Reliable, Traceable

© 2022 CGI Inc.

Thinking to the Future



Bringing together of experts

Establish building blocks

Scientific/Technology Roadmap

Based on prototyping activities

- Integration of models from different domains – further work to be done
- Bring in data sources from concurrent digital twins like the precursor projects
- Ensure they do not end up in silos – boundary conditions between them next big issue to address
- Help ESA and EU's vision to achieving Digital Twin Earths to help humanity have a sustainable future by providing invaluable insights we can act on

CGI esa

OXFORD UNIVERSITY INNOVATION

UNIVERSITY OF OXFORD

TRILLIUM CLIMATE | OCEAN | SPACE

IIASA

Forest DTEP
Part of ESA's Digital Twin Earth

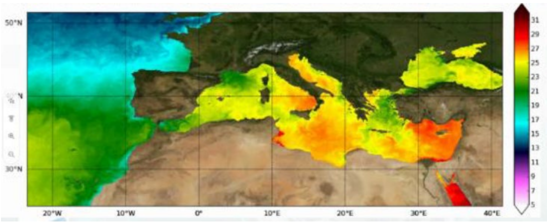
Digital Twin Antarctica

CLIMATE IMPACT EXPLORER
DIGITAL TWIN EARTH PRECURSOR

DTE Hydrology

Oceans

Ifremer



Food Systems