Belmont Forum

UNIVERSITY OF OXFORD

E-INFRASTRUCTURES & DATA MANAGEMENT
Collaborative Research Action

GOTHAM – Globally Observed Teleconnections in Hierarchies of Atmospheric Models

David Wallom, University of Oxford e-Research Centre

Scoping Workshop

November 28-29, 2016

ANR, Paris

ANR AGENCE NATIONALE DE LA RECHERCHE

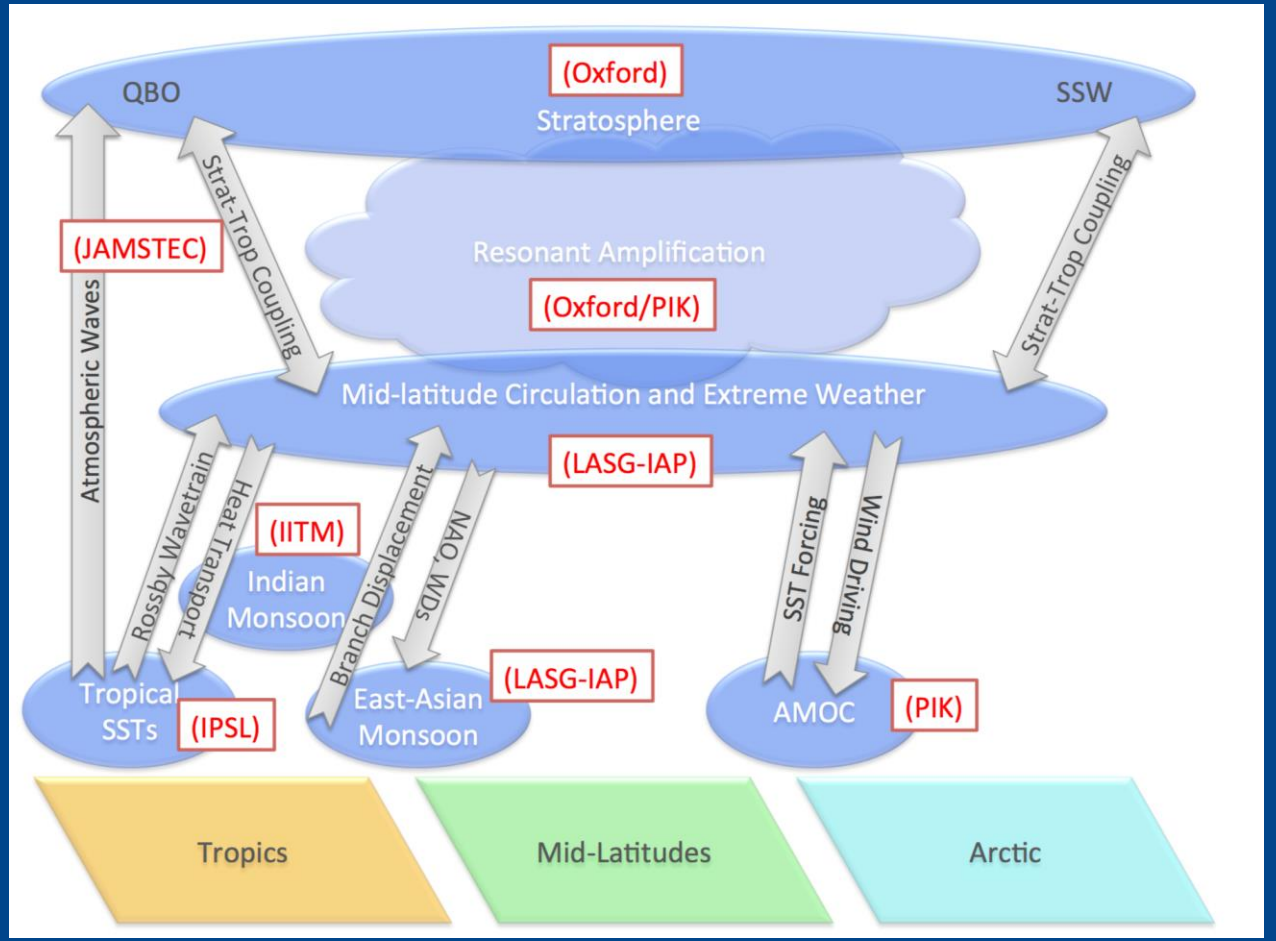
BELMONT FORUM

GOTHAM PROJECT DESCRIPTION

- 4 year Belmont Forum/JPI Climate funded consortium,
  - Investigator groups from: India (**IITM**), China (**LASG-IAP**), Japan (**JAMSTEC**), United Kingdom (**Oxford**), France (**IPSL**), Germany (**PIK**);
  - Project partners from Australia (**CAWCR**), Canada (**Environment Canada**), USA (**NCAR**), Finland (**FMI**), Denmark (**DMI**) and United Kingdom (**Met Office**, **CEH**, **NCAS**)
- Investigating teleconnections between the tropics and high latitudes and extremes impacting vulnerable areas outlined in latest IPCC Report
- Methodologies include: bespoke experiments with investigator group GCMs, massive ensembles $O(10^5)$ produced by distributed computing (climateprediction.net & weather@home), Project partner submissions to **SPARC-QBOi** experiments and advanced mathematical techniques (Complex Climate Networks)

GOTHAM INSTITUTIONAL INTERESTS & SYNERGIES

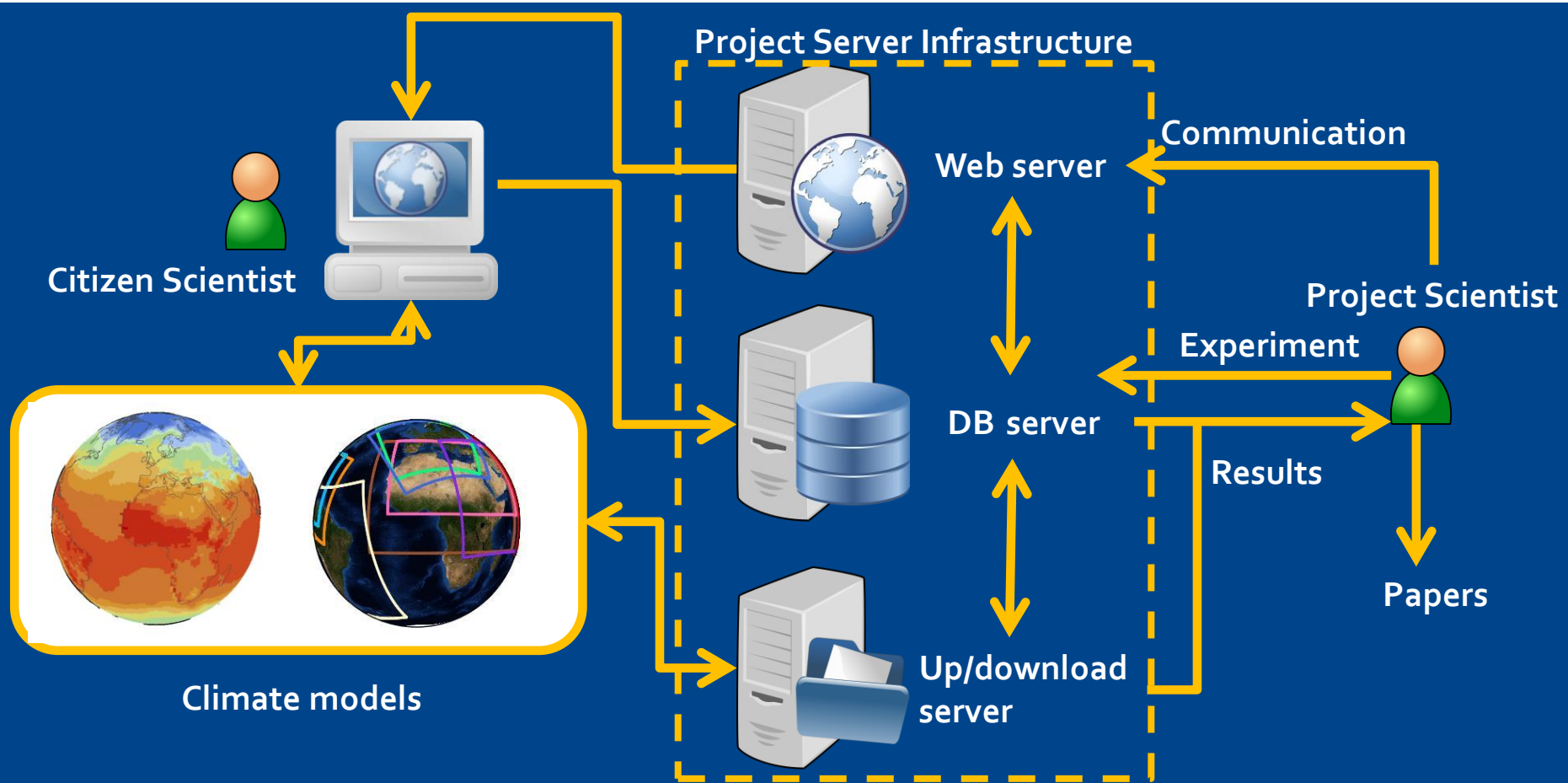
- How well do current seasonal forecast and climate prediction models capture tropical-polar teleconnections, including those via the stratosphere? (Gray, Osprey)
- Which global teleconnection patterns most influence regional climate and extreme weather within the tropics and subtropics, in particular the Asian Monsoon? (Wu)
- How predictable are extreme El Nino events and how diverse are the regional responses to them? (Lott, Guilyardi)
- Which feedbacks and relevant processes contribute to maintain mid-latitude circulation regimes? (Coumou, Donner)
- How do the Asian monsoon and mid-latitude circulation interact leading to extreme events? (Krishnan, Vellore)
- How does stratospheric circulation impact patterns of tropical precipitation? (Watanabe, Kawatani)



GOTHAM E-INFRASTRUCTURES & DATA MANAGEMENT

- e-Infrastructures
  - Computation
    - HPC
      - National scale facilities utilised within all workpackages for high resolution simulation
    - Volunteer Computing
      - Utilised within two work packages to provide massive ensembles to allow statistical analysis
    - Public Cloud
      - Availability of public cloud resource to support urgent 'time to solution' problems
  - Data
    - Management
      - Project Data Director and Data Manager appointed
      - Individual workpackage data managers to be appointed
      - Project data management plan created as living document
        - Created to satisfy individual national funding body requirements on long term accessibility of data
    - Single project repository for data through JASMIN Group workspace (100TB)
      - Active and semi-active data, i.e. data actively being generated and used in active publication prep
      - To fully support responsibilities we need clarity on support of non-NERC funded storage for longer term
    - Formats
      - Its not just about final output data formats!
      - Coherent ancillary data across modelling groups where possible to allow inter-comparison of results
  - People
    - Every researcher needs to be a digital scientist, there cannot be 'domain separate from digital'!
    - Annual Summer schools with contributions

CLIMTEPREDICTION.NET & WEATHER@HOME - VOLUNTEER DISTRIBUTED COMPUTING



Very large ensembles of simulations can be generated by using this framework.

EXPECTATIONS FROM E-I&DM CALL

e-Infrastructure

- MUST NOT FUND/SUPPORT JUST 'MORE OF THE SAME', THERE IS A LOT OF BIG SUPERCOMPUTERS!
- Open standards in all areas allow for interoperability and interoperation, technically and socially
- Computation; [HPC, Cloud, Volunteer Computing] all are useful e-infra for different types of activities and scales of work, we must ensure that any call respects that one size will not fit all and not prejudge what it is going to fund.
- Must be able to move computation to data, data must be globally accessible, not just the 1st world resource rich academic community
- Data/Metadata - desirable to support a single responsible actor/organisation across all Belmont projects who will either;
  - receive data in a number of agreed formats and translate into a single coherent dataset
    Or
  - Agreed set of data formats from different participant community

  for longer term archiving and possible reuse.
- [Need to consider long term value as regeneration may be better and more responsible]
- Software – Co-design of applications with new hardware is the only route to maximise performance, rather than just hoping that hardware speedup will solve our problems
- People - Every researcher must be interdisciplinary, every researcher must be a computational scientist from their first exposure as undergraduates to the end of their career as senior PI
- People – There must be a democratisation where every researcher from every country is equal, there is the recognition that there is need for a coherent community, all are equally skilled , ensuring equal contribution.
- People – Engagement with non-academic commercial partners for impact studies – Recognition of service quality and SLA requirements