

CMIP6 Infrastructure

Karl E. Taylor & V. Balaji

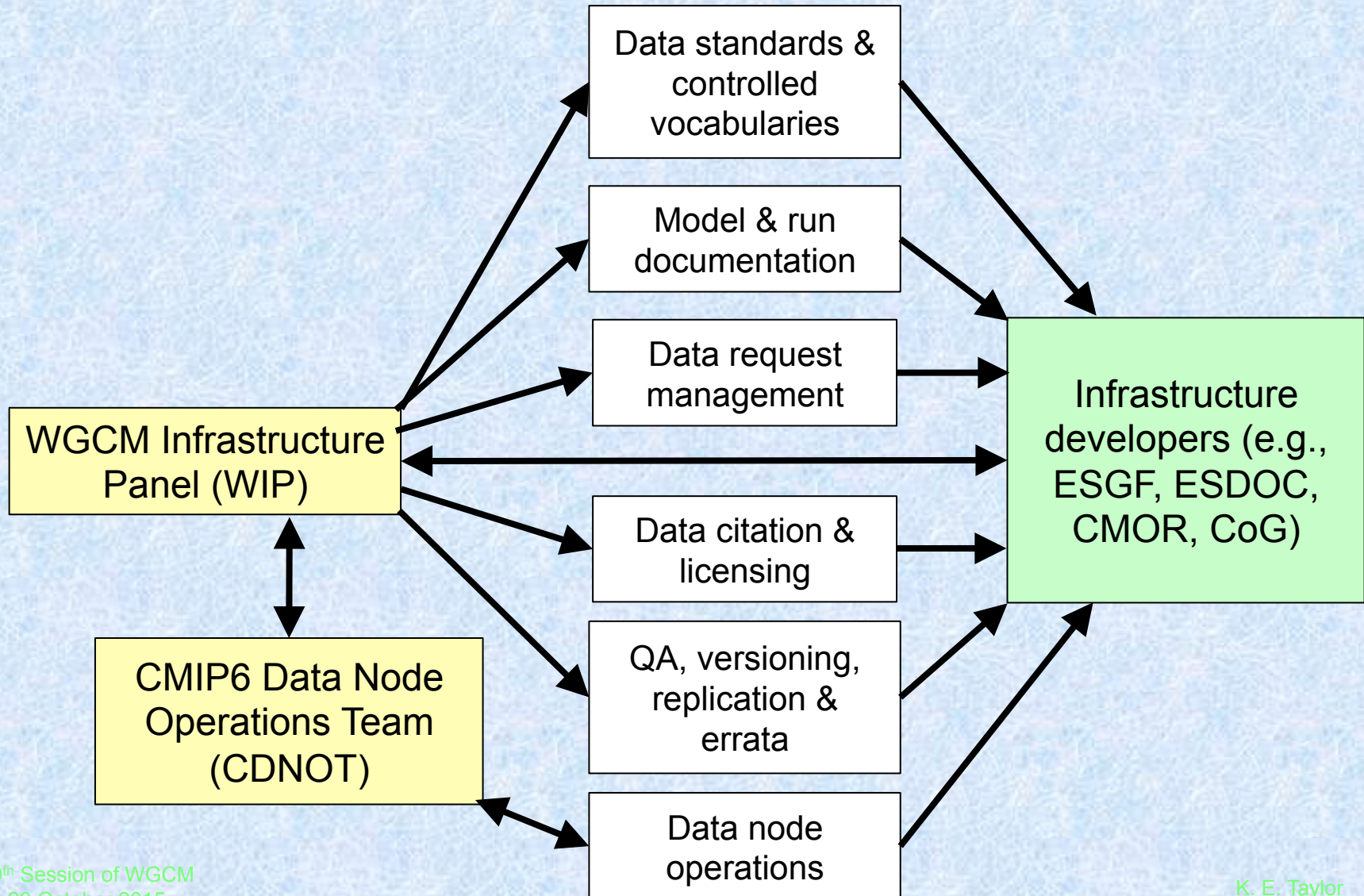
Presented at Workshop on
CMIP5 Model Analysis and Scientific Plans for CMIP6

Dubrovnik, Croatia
20-23 October 2015

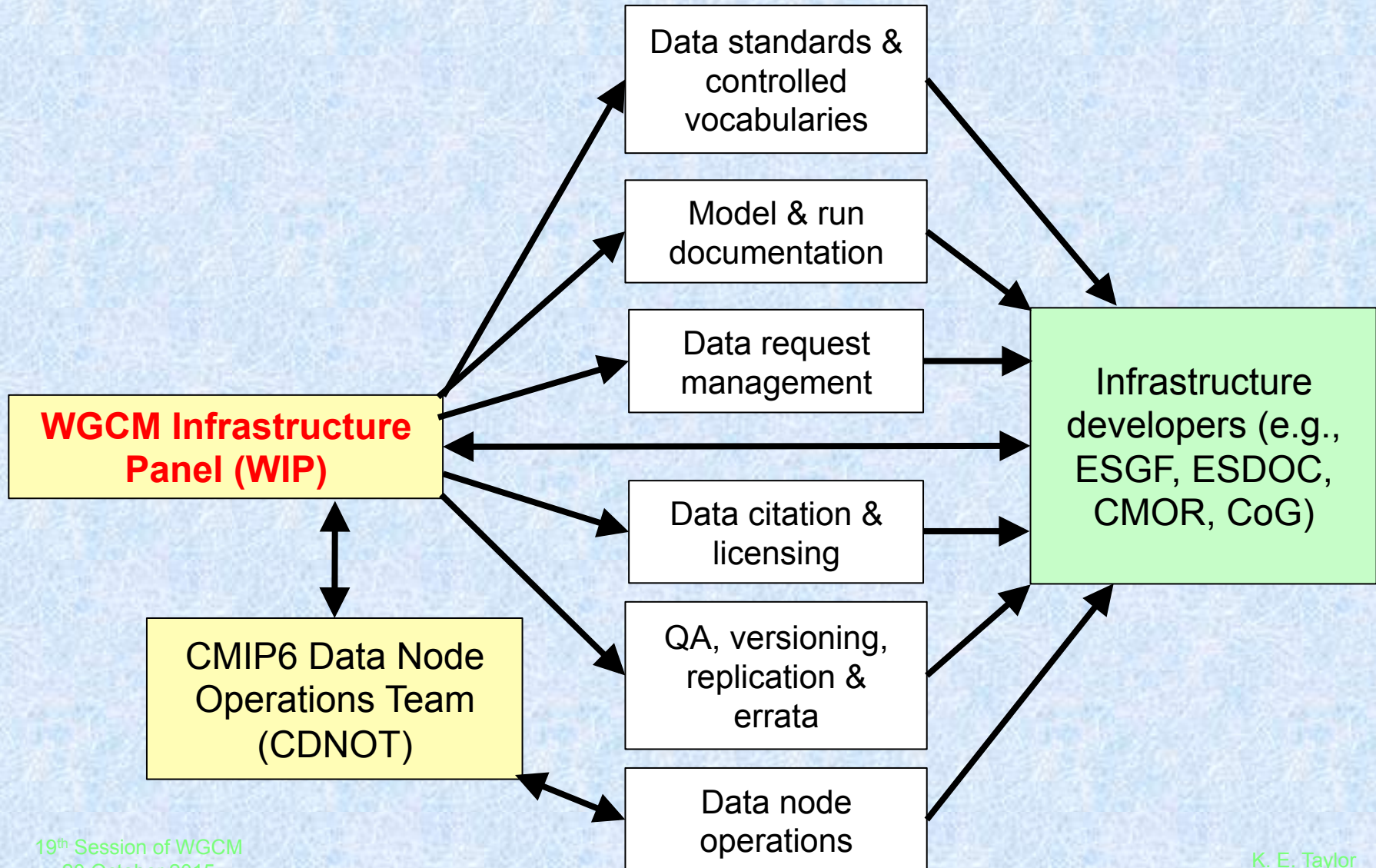
CMIP coordination

- The WGCM oversees and provides direction
- The CMIP Panel manages and coordinates scientific aspects (experiment design, output request, forcing data sets, etc.)
- The WGCM Infrastructure Panel (WIP) manages and coordinates infrastructure development, implementation, and operations.

MIP infrastructure coordination, development, and implementation



MIP infrastructure coordination, development, and implementation



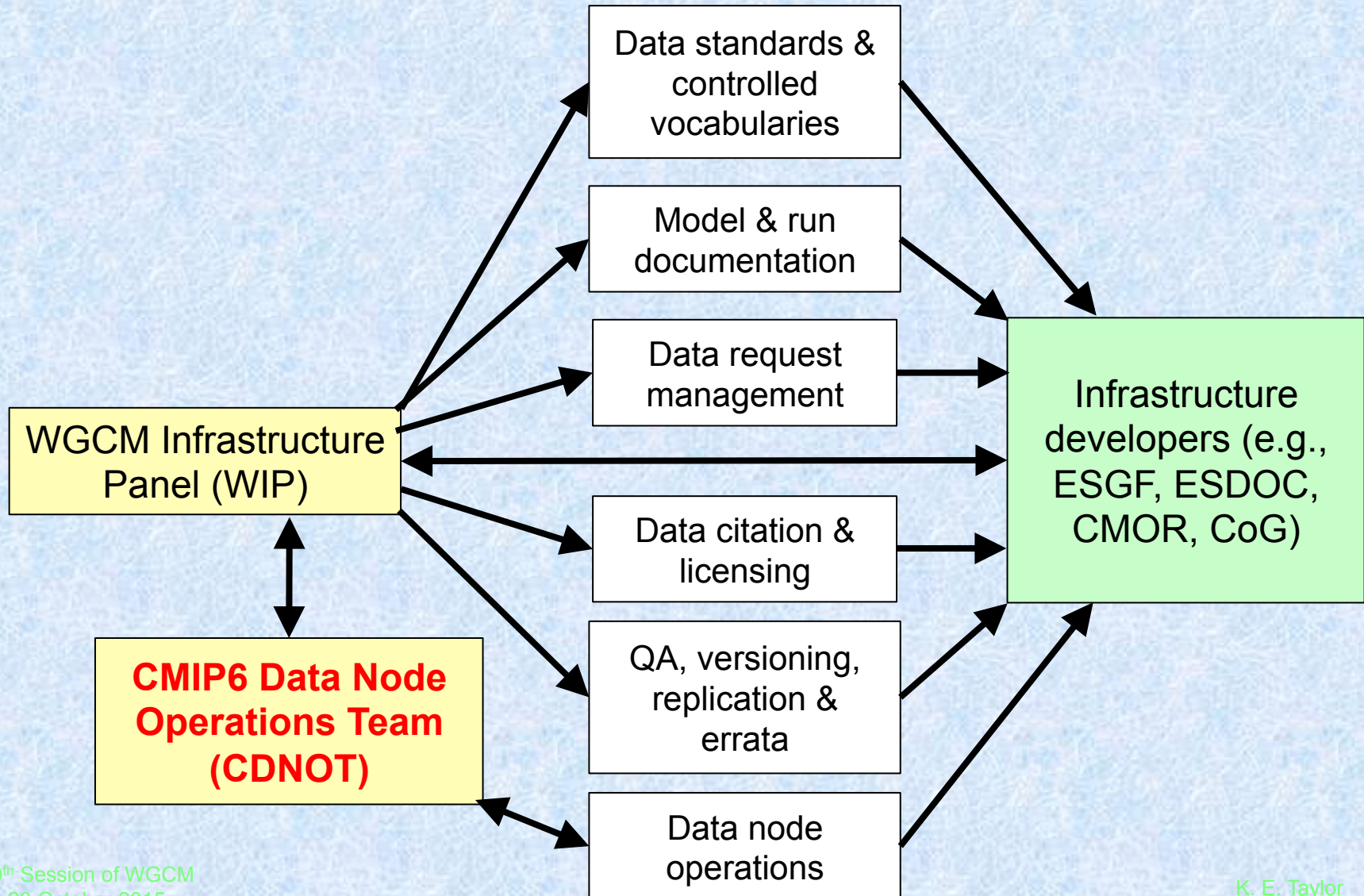
Why a WCRP Infrastructure Panel (WIP)?

- CMIP and other “MIPs” represent a huge investment by modeling groups
- MIPs provide an enormously valuable resource of simulation output for scrutiny by scientists
- The CMIP panel attempts to coordinate MIP design to maximize scientific return on the resource investment
- The WIP coordinates the technical aspects of MIPs to reduce the burden on modeling groups and facilitate analysis.

WIP members: computer and climate scientists representing data centers and modeling groups

V. Balaji (co-chair): GFDL
Karl Taylor (co-chair): PCMDI
Luca Cinquini: NASA JPL
Cecelia DeLuca: NOAA
Sébastien Denvil: IPSL
Mark Elkington: MOHC
Francesca Guglielmo, LSCE
Eric Guilyardi: IPSL
Martin Juckes: BADC
Slava Kharin: CCCma
Michael Lautenschlager: DKRZ
Bryan Lawrence : NCAS, BADC
Dean Williams: PCMDI

MIP infrastructure coordination, development, and implementation



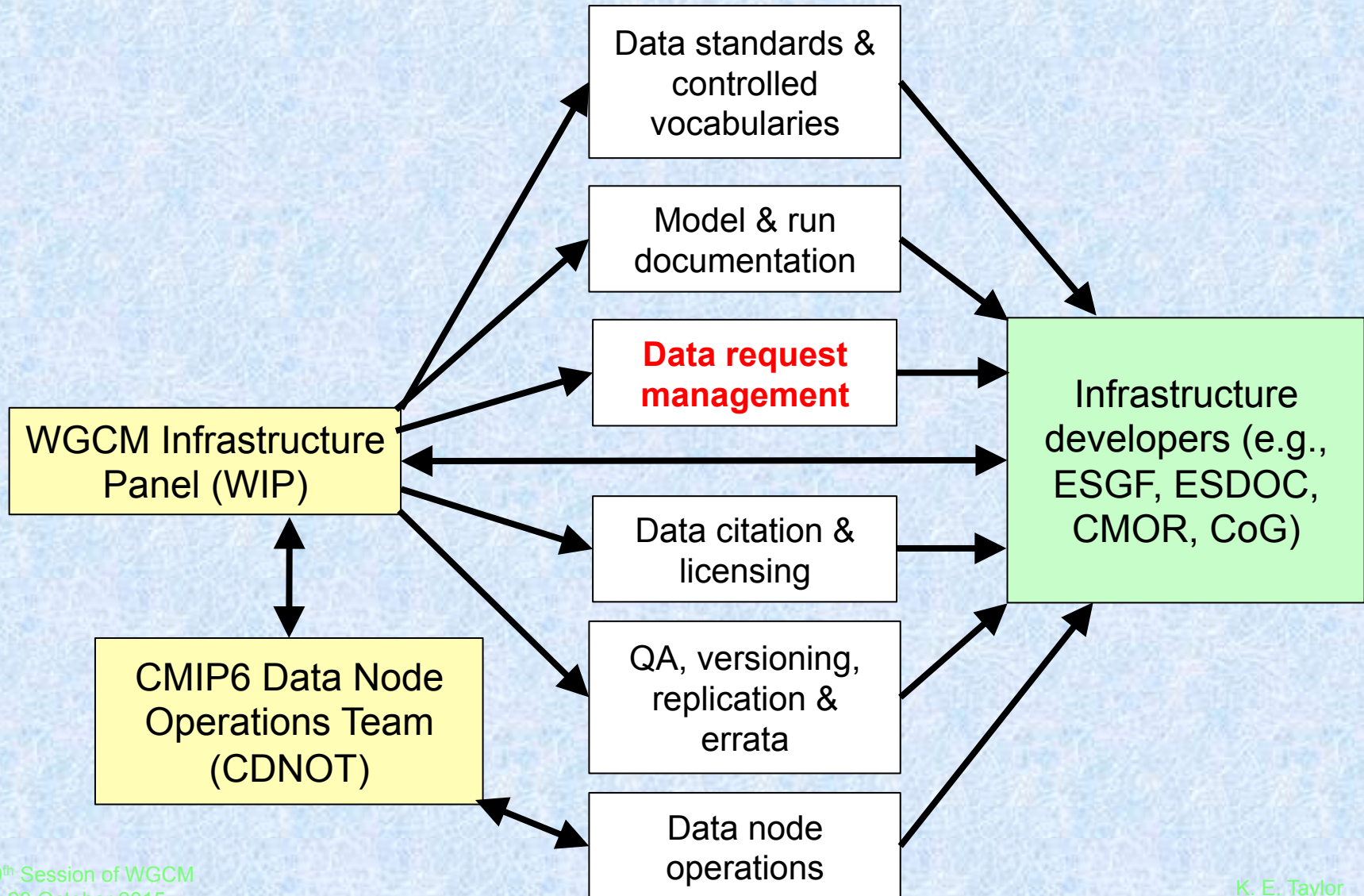
CMIP Data Node Operations Team (CDNOT)

- A technical consortium charged with operationalizing the CMIP6 ESGF Federation
- Sébastien Denvil (IPSL) chairs
- Members representing each site hosting CMIP6 data (i.e., most modeling centers and major data centers)
- Membership overlaps with bodies responsible for requirements (WIP) and software development (ESGF, ESDOC, ...)
- Serves to:
 - Communicate WIP discussion to all those of interest
 - Provide input to the WIP of data node/modeling center concerns

WIP work

- Write position papers detailing
 - CMIP requirements for infrastructure
 - Technical specifications for all aspects of model output
- 10 papers in the works (half ready for community review)
- Poll MIPs for input on model output needed
 - Write document providing guidance on data request
 - Write code translating MIP input into xml's
- Work toward consensus on:
 - Regridding data to commons grid(s)?
 - Compression of data

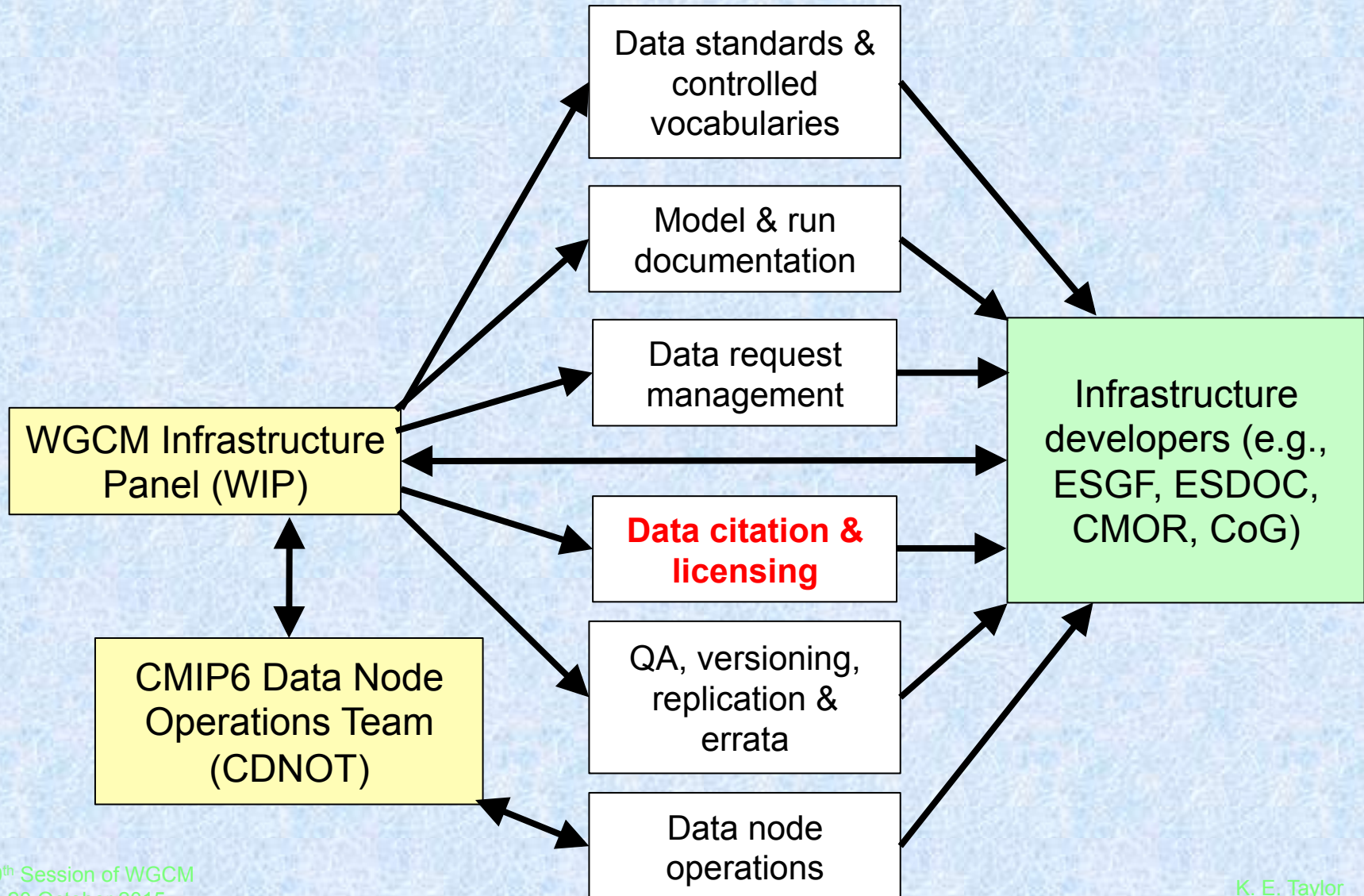
MIP infrastructure coordination, development, and implementation



4 Position papers: CMIP6 data request management

- Led by Martin Juckes (STFC)
- Documents dealing with technical aspects of CMIP6 data:
 - Compilation of list of variables
 - Specification of file name template and file global attributes
 - Specification of “controlled vocabulary” enabling automated management and utilization of CMIP archive
- Status:
 - MIPs submitted list of variables they need
 - CMIP Panel will review for completeness and reasonableness
 - Iterate with MIPs and modeling groups

MIP infrastructure coordination, development, and implementation

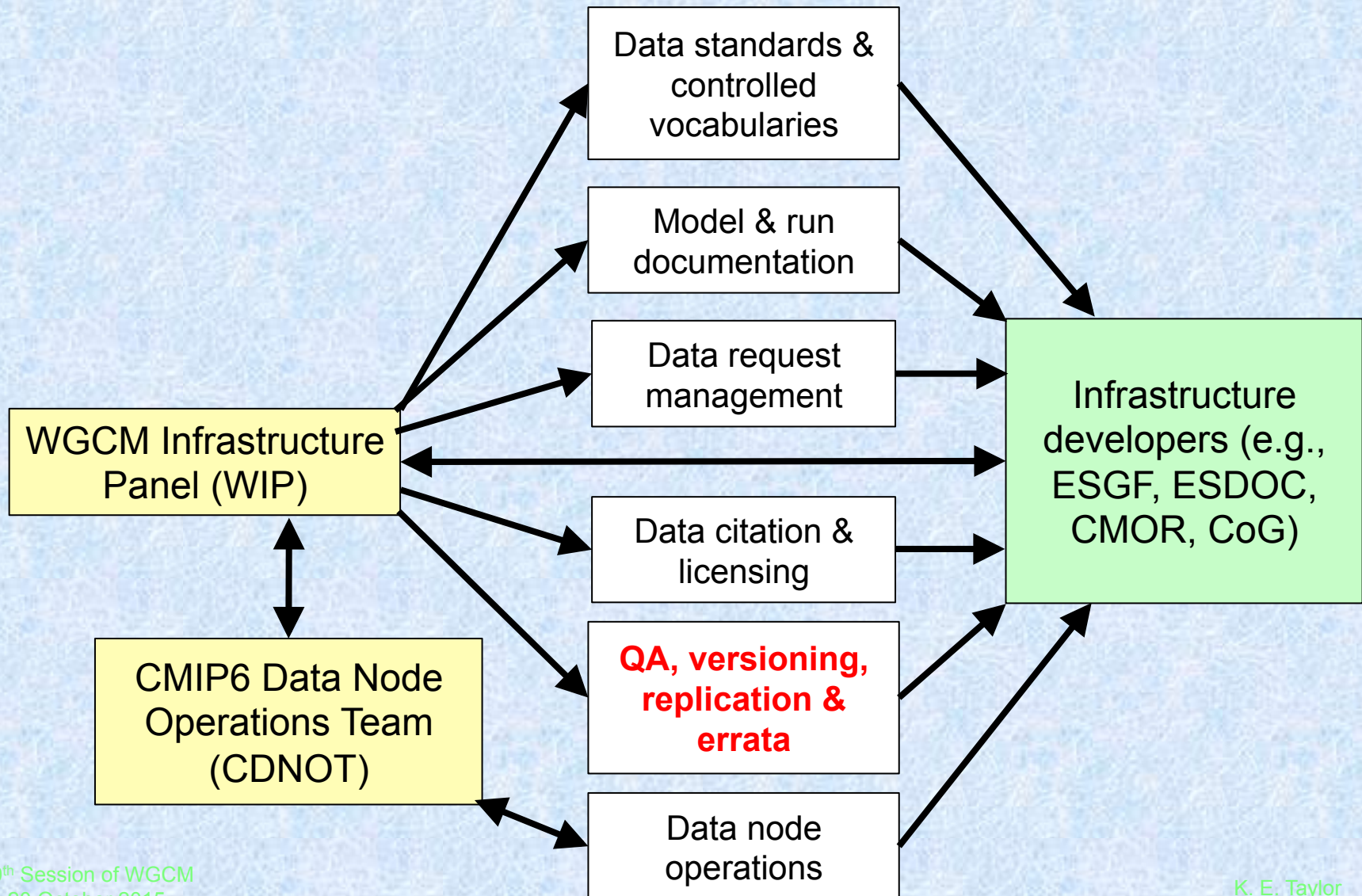


2 Position papers: Data citation and licensing

Main requirement: ensure proper citation of data acknowledging contributions by modeling groups and simplified access control on ESGF

- DOI generation at the granularity of model and simulation.
- Citation included in the terms of use of CMIP6 model output.
- Links connecting datasets to model and experiment documentation (ESDOC/CIM)
- Recording of Creative Commons Licensing agreements in files.

MIP infrastructure coordination, development, and implementation



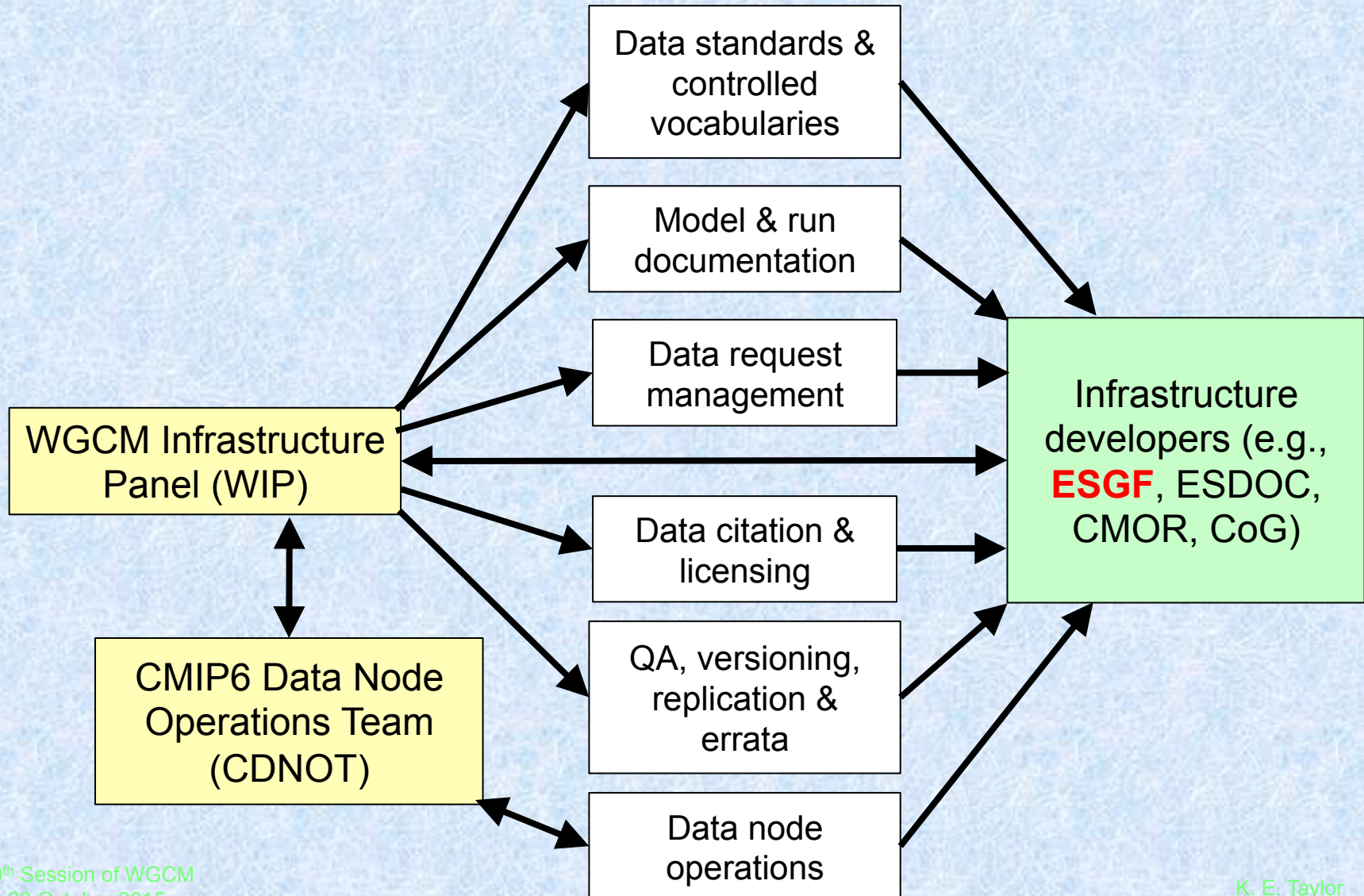
4 Position papers: Quality assurance (QA), replication, versioning, and errata

Main requirement is for end users to know if they are working with the right dataset in a federation where data is replicated multiple times, may have been retracted or superseded.

Highlights:

- Extend use of persistent identifiers (PIDs) for dataset tracking (replaces **tracking_id** from CMIP5).
- Lists of PIDs can be used as supplementary citation information in papers.
- PID-based query system to see if errata have been reported, or data have been superseded.
- Potential use of PIDs in replication workflow.
- Automated QC mechanisms to ensure adherence to metadata and data quality standards.

MIP infrastructure coordination, development, and implementation



ESGF status

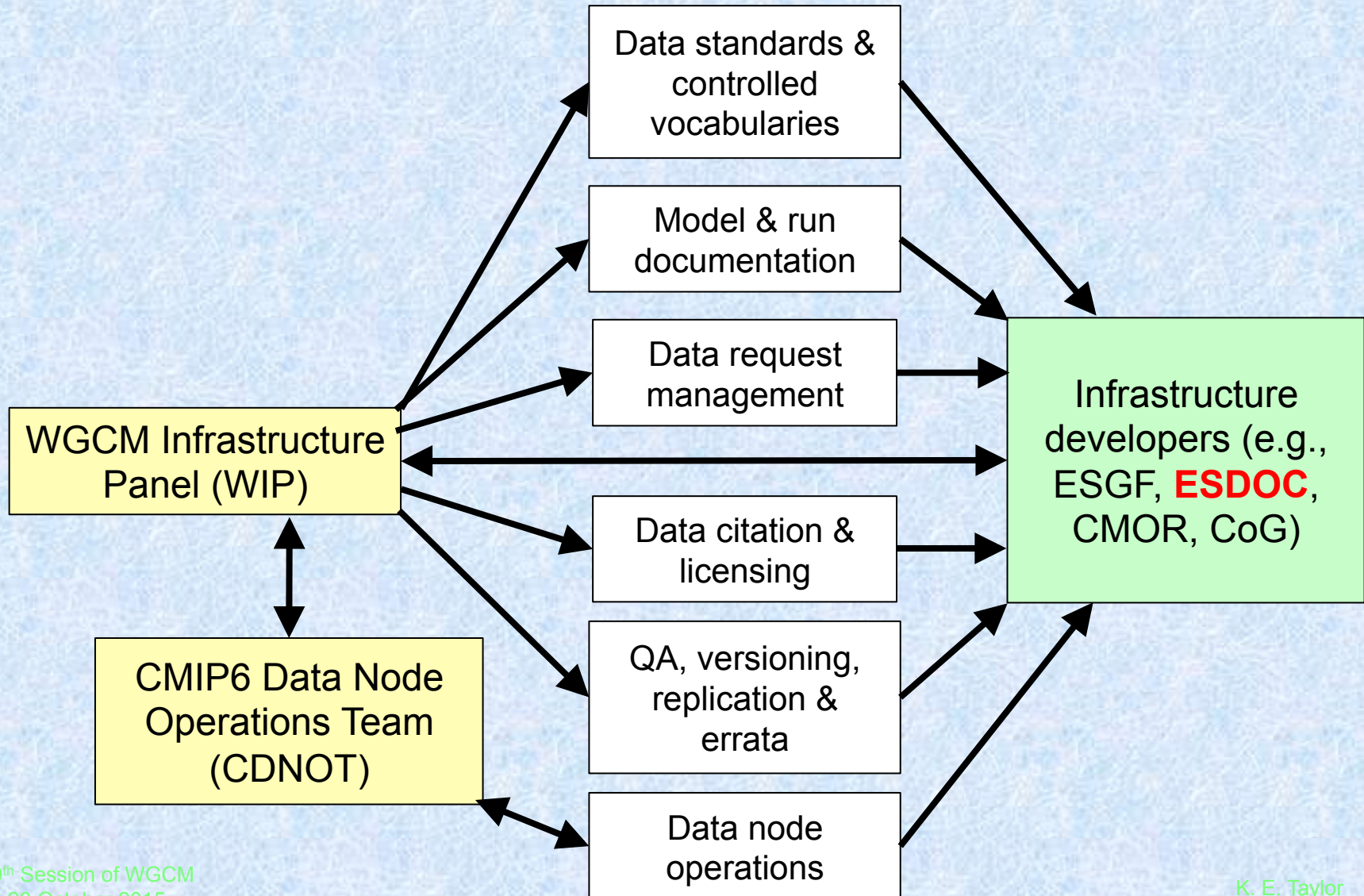
Following a security incident in June 2015, the ESGF system was brought offline and the software stack was extensively re-engineered to accomplish the following goals:

- Execute complete software scan of all modules, fix all exposed and other potential security breaches
- Major upgrade of underlying system infrastructure
- **CMIP5 should be back online the first week of November**

Major ESGF upgrades

- Upgrade underlying system infrastructure:
 - CentOS7, Java 1.8, Tomcat 8, Postgres 9.4, OpenSSL 1.0, Python 2.7.9
 - Switch ESGF installer to RPM-based components
 - Run Apache httpd server in front of Tomcat (better performance, flexibility)
- Upgrade of all ESGF services:
 - Search services (Solr5), data download (TDS5), high performance data transfer (Globus-Connect-Server), computation (UV-CDAT), visualization (LAS)
 - Replace old web-front-end with new CoG user interface
- Republish ALL data collections (CMIP5, CORDEX, Obs4MIPs, ana4MIPs,...)

MIP infrastructure coordination, development, and implementation



ES-Doc: Model & simulation documentation

- Model documentation is essential for interpretation of results.
- CMIP5 viewer & comparator operational
(see <http://es-doc.org> -- click on "tools")
- CMIP6 planning:
 - Simplification of contents relative to CMIP5
 - Better documentation of forcing.
 - Quality Control Information (of the simulations and of the descriptions)

Conclusions

- WGCM Infrastructure Panel translates CMIP experimental design into requirements for the global data infrastructure
- Governance at different stages of infrastructure: requirements (WIP), software development (ESGF, ESDOC, CMOR, ...), CMIP6 implementation and operations (CDNOT).
- Close involvement of WIP with ESGF-XC and CDNOT (overlapping membership)
- WIP has produced 11 position papers available soon on the WIP website.

We invite input.

