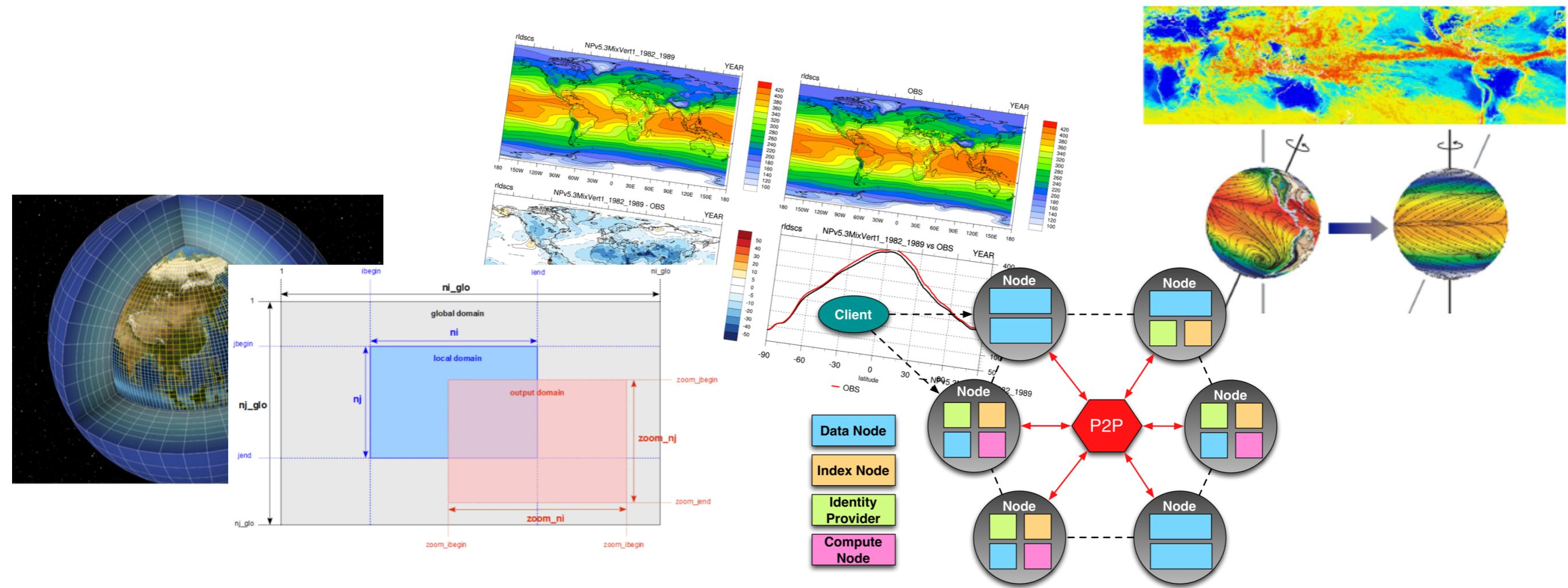


# The Synda software

Jérôme Raciazek (IPSL), Guillaume Levavasseur (IPSL), Sébastien Denvil (IPSL).



# Synda & the user survey

Question	Category	1	2	3	4	5	6	Total answers	Weighted Scores	% of Responses
Which feature do you find most difficult to use ?	Globus download	3	4	6	4	12	17	46	4,50	13,41%
Which features of ESGF do you find most useful?	Globus download	3	7	7	3	7	7	34	3,74	9,91%
Which feature do you find most difficult to use?	LAS visualization engine	4	4	5	7	7	13	40	4,20	11,66%
Which features of ESGF do you find most useful?	LAS visualization engine	6	6	7	7	4	7	37	3,49	10,79%
Which feature do you find most difficult to use?	Synda download client	10	5	2	3	5	9	34	3,44	9,91%
Which features of ESGF do you find most useful?	Synda download client	4	1	5	3	6	10	29	4,24	8,45%

# Synda acronyms

**SDT** : SYNDA DATA TRANSFERT

**SDP** : SYNDA DATA PROCESSING

**SDW** : SYNDA DATA WORKER



# Synda in a nutshell

BASIC  
USAGE

- Command-line
- ESGF mirroring
- Parameter discovery
- Data discovery
- Similar to “apt-get”  
but asynchronous

ADVANCED  
USAGE

- Data management
- Use nearest replica
- GridFTP
- Post-processing
- Downloading supervision



<https://github.com/Prodiguer/synda>

# Synda in more details

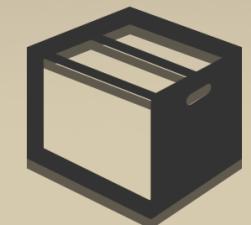
- RPM & DOKER package
- install.sh script.

## INSTALLATION PROCEDURE



- Daemon mode running in background,
- Online help.

## SYSTEM PACKAGE



- **Support all ESGF project** (CMIP5, CORDEX, etc.),
- **Facetted data selection** (experiment, variable, etc.),
- **Incremental search** (download only what's new),
- **Nearest replica selection**.

## SEARCH



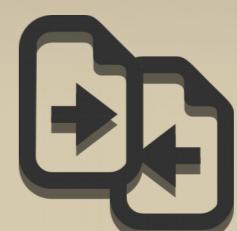
- **Parallel downloads** (using HTTP or GridFTP protocol),
- Checksums control,
- Transfer priority and statistics,
- **Transparent x509 certificate renewal**.

## DOWNLOAD

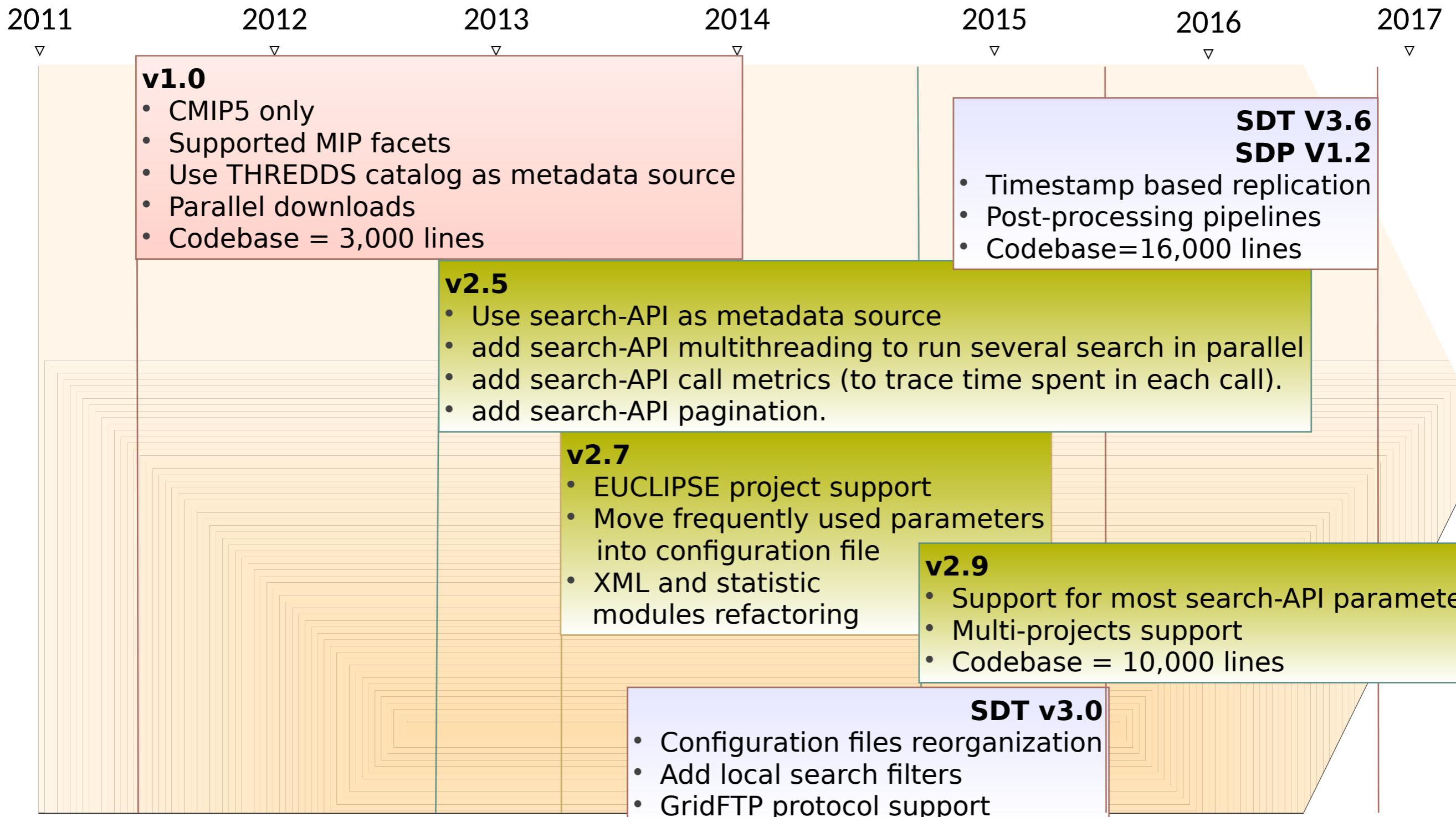


- **Default storage following DRS tree format**,
- **Template based data management**,
- **Auto-remove old dataset versions**.

## DATA MANAGEMENT



# Synda history



The program evolve together with the ESGF archive backend functionalities.  
 5 releases this year (2016)

# Global data infrastructure

How long does it take you on average to discover and access the date and ressources you need ?

**Table 8. How long does it take you on average to discover and access the data and resources you need?**

	Minutes		Hours		Days		Can't Find/Access		Total	Weighted Average
<b>Discover</b>	49.53%	105	36.79%	78	10.38%	22	3.30%	7	<b>212</b>	<b>1.67</b>
<b>Access/Download</b>	12.92%	27	42.11%	88	41.63%	87	3.35%	7	<b>209</b>	<b>2.35</b>

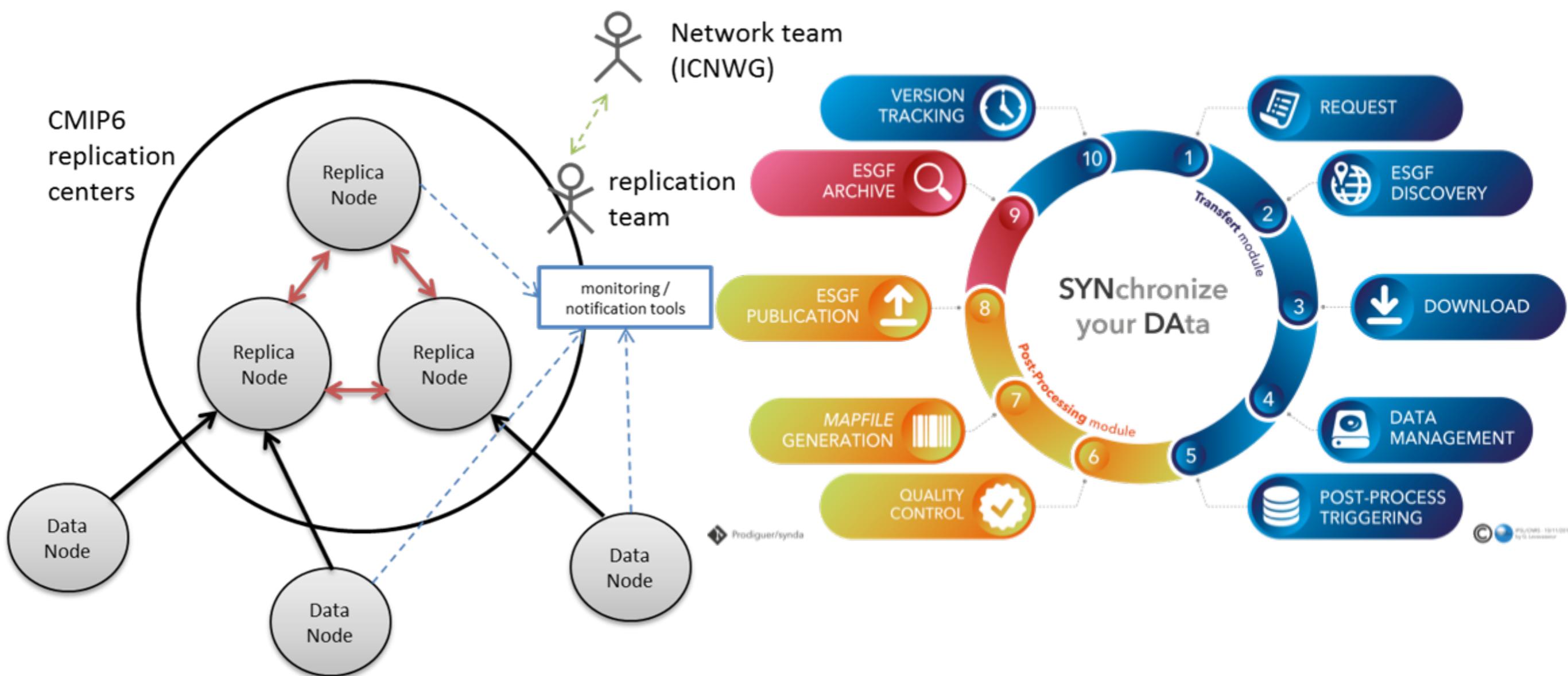
Which takes the longest to discover and use ?

**Table 9. Which takes the longest to discover and use?**

	1 (Shortest)		2		3		4 (Longest)		Total	Weighted Average
<b>Data</b>	16.96%	29	22.22%	38	25.15%	43	35.67%	61	<b>171</b>	<b>2.80</b>
<b>Computer</b>	31.01%	49	49.37%	78	14.56%	23	5.06%	8	<b>158</b>	<b>1.94</b>
<b>Storage</b>	21.52%	34	36.08%	57	25.95%	41	16.46%	26	<b>158</b>	<b>2.37</b>
<b>Software</b>	24.53%	39	34.59%	55	20.75%	33	20.13%	32	<b>159</b>	<b>2.36</b>

# Replication & versioning

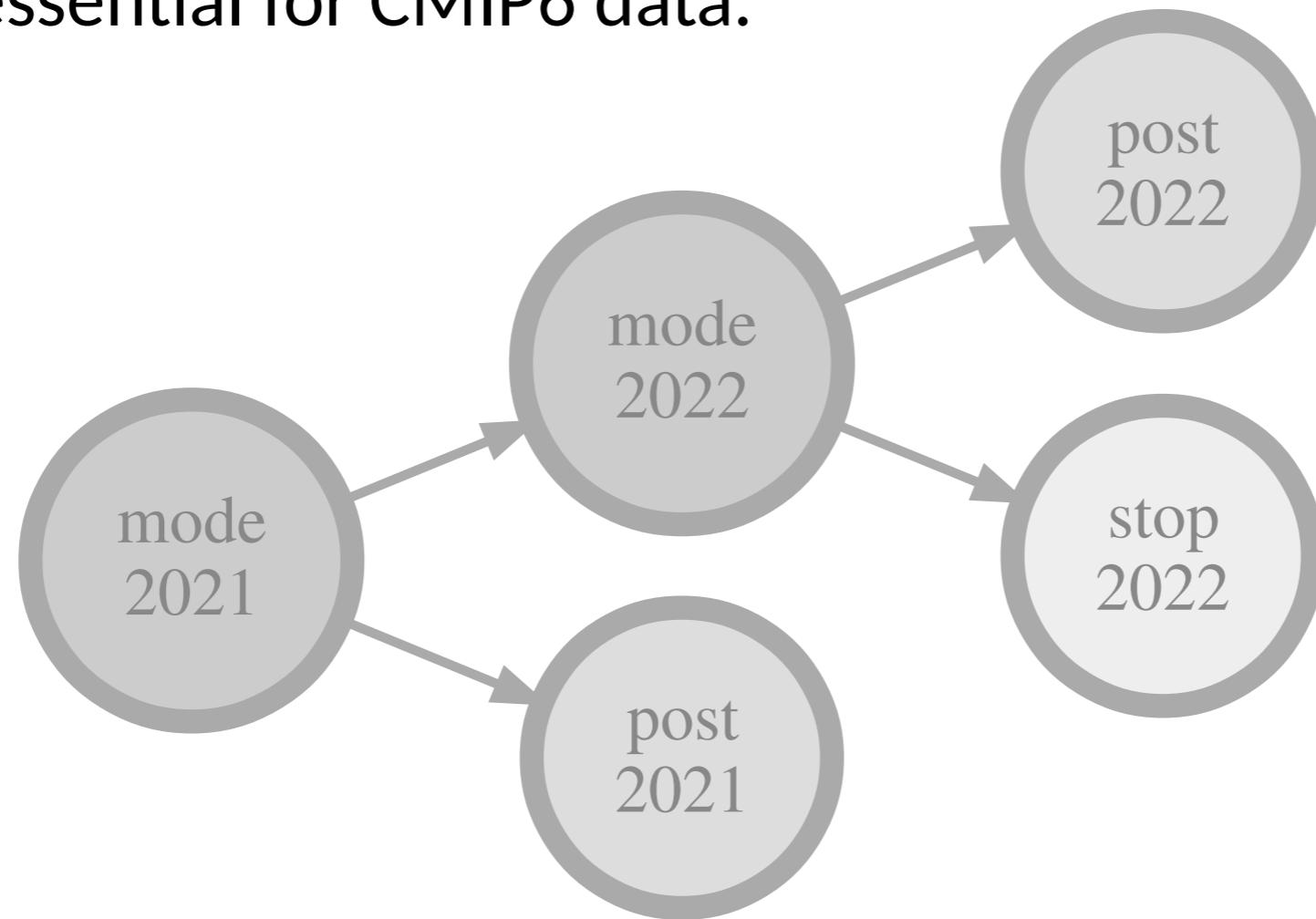
- Impact on CMIP6 data management (DM) and ESGF governance (ESGF)
- Stable processes which are supervised by a board (the CDNOT Team?) are needed for CMIP6 data consistency in ESGF
- CMIP6 data replication architecture:



# Synda main achievement

CMIP6 horizon: increasing data amount and data management

No fast replica synchronization or notification mechanisms are currently supported. Thereby in general original data can be unpublished or modified without effects on replica sites. Automatic replication procedures are considered essential for CMIP6 data.



# Synda state of play

- Improvements for admin usage:
  - SDT allows apart selection files, default files.
  - SDP allows pipeline management and dependency out of the box
- Finer DRS management
  - « custom\_dataset\_template » can be used in SDT default files to stick to the official directory structure of a project
  - I/O directories improved for SDP through flexible JSON files and new arguments for SDW (the worker)
- ESGF replication
  - SDP ready for production phase.
  - Mapfile and publication workers are available in agreement with the last publisher release
  - Other QC workers can be added to the replication pipeline if desired
    - Time axis and period overlap checking (available)
    - QA-DKRZ (worker under active development)



# Thank you for your attention

