

# Οι υπηρεσίες ΤΠΕ στην υποστήριξη της έρευνας στο ΑΠΘ

## *Προοπτικές Open Data & Open Science*



*Π. Κορόσογλου*

*Κέντρο Ηλεκτρονικής Διακυβέρνησης ΑΠΘ*

*Γρ. Επιστημονικών Υπολογιστικών Υπηρεσιών*

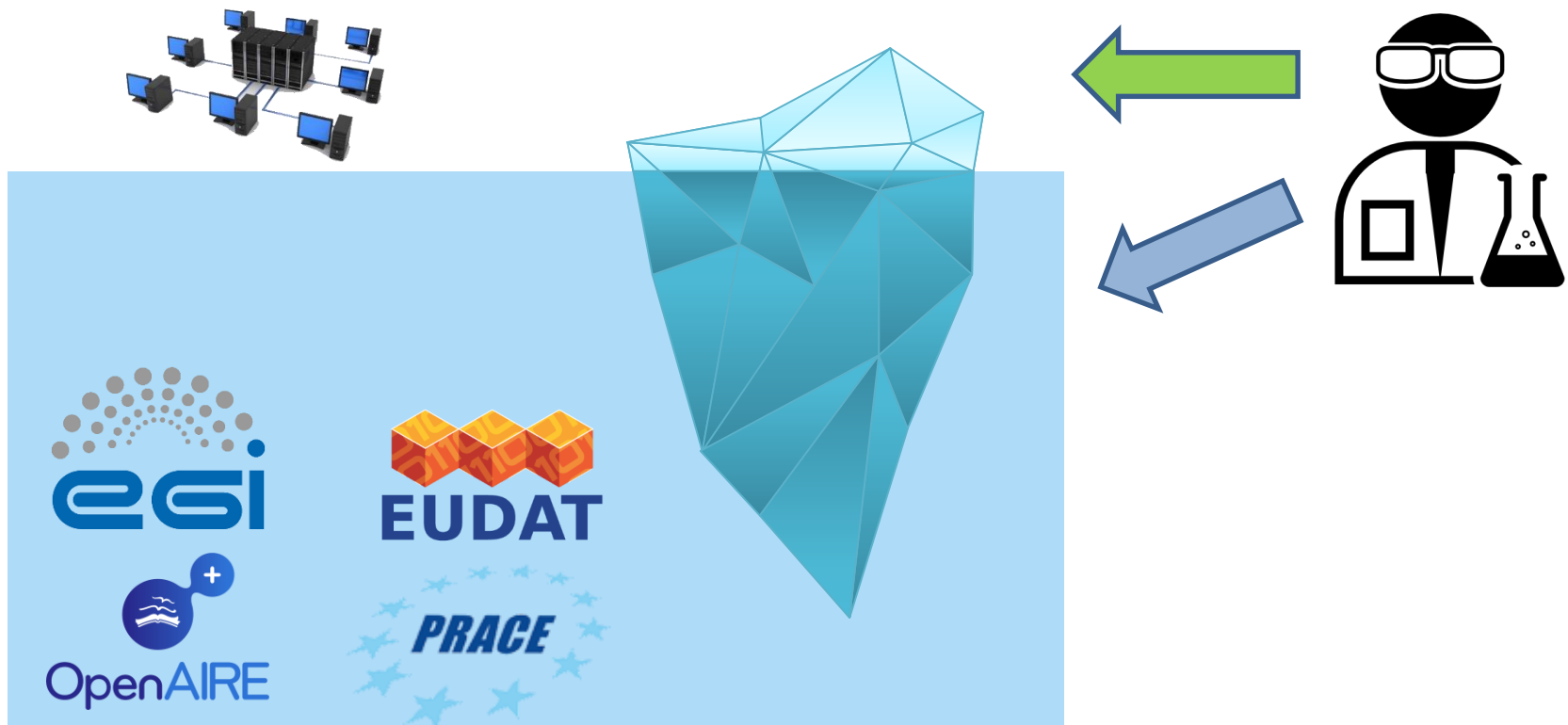
# Παρατήρηση/Πείραμα/Προσομοίωση

- Η παρατήρηση οδηγεί στην διατύπωση της θεωρίας
- Το πείραμα επιβεβαιώνει ή όχι τη θεωρία
  - Την επιβεβαιώνει εφόσον είναι αναπαράξιμο
- Η προσομοίωση αξιολογείται με βάση τη διατύπωση και παράγει νέα αποτελέσματα (δεδομένα) ερευνητικά
  - Κατά πόσο όμως είναι διαθέσιμη, αναπαράξιμη και πώς;
  - Τα ερευνητικά δεδομένα εισόδου ή εξόδου είναι διαθέσιμα για νέα χρήση ή μετα-ανάλυση;
  - Ποιό το κέρδος αν υπάρχει για τον/-ην ερευνητή;

# Outline

- Υποδομές: τοπικές (ΑΠΘ) & απομακρυσμένες
- Υπηρεσίες ΤΠΕ στο ΑΠΘ για την έρευνα
  - Computing
  - Storage
  - Big Data
  - Open Data (σημερινή κατάσταση & προοπτικές)
  - Open Science (προοπτικές)

# Τοπικές/Διεθνείς υποδομές



Taken and adapted with the kind permission of Dr. Fotis Psomopoulos. Original slides:  
<https://www.digitalinfrastructures.eu/content/e-infrastructures-bioinformatics-long-tail-science-user-perspective>



- Grid
  - 325 Grid Sites
    - 826,500 CPU Cores
    - >500PB Storage
    - ~1.5M Jobs
- Federated Cloud
  - 21 Cloud Sites
    - 6600 CPU Cores
    - ~100VMs/day
- Usage Metrics
  - 46K users
  - >400 research papers in 2015



- 25 Member Countries
- 7 Tier-0 offered by 5 PRACE members
  - CURIE (@GENCI, France) => **1.75PFlops** & 15PB
  - MARCONI (@CINECA, Italy) => **13PFlops**
  - Hazel Hen (@HLRS, Germany) => **7.42PFlops** & 11PB
  - JUQUEEN (@FZJ, Germany) => **5.87PFlops**
  - MareNostrum (@BSC, Spain) => **1PFlops**
  - Piz Daint (@CSCS, Switzerland) => **7.8PFlops**
  - SuperMUC (@LRZ, Germany) => **3.2PFlops** & 10PB



- Collaborative Data Infrastructure
  - 35 European Organizations from 15 Countries
- Services:
  - Handle service (i.e. for data registration)
  - Search service (you know, for Search)
  - Staging service (i.e. for computation)
  - Replication service (i.e. for safekeeping)
  - Storing and sharing services
  - Syncing service



- Open Access integration service
  - Aggregates de-centralized publications and/or data repositories
  - Central Zenodo service @CERN
- Content providers
  - Journals
  - Institutional repositories
  - Data repositories
  - Publication repositories
  - Thematic repositories



# Υποδομές ΤΠΕ στο ΑΠΘ για έρευνα

12 Rack

>200 Physical Hosts

>50 Virtual machines

>30 Switches

4 A/C units

1 + 4 UPS units

1 Power Generator

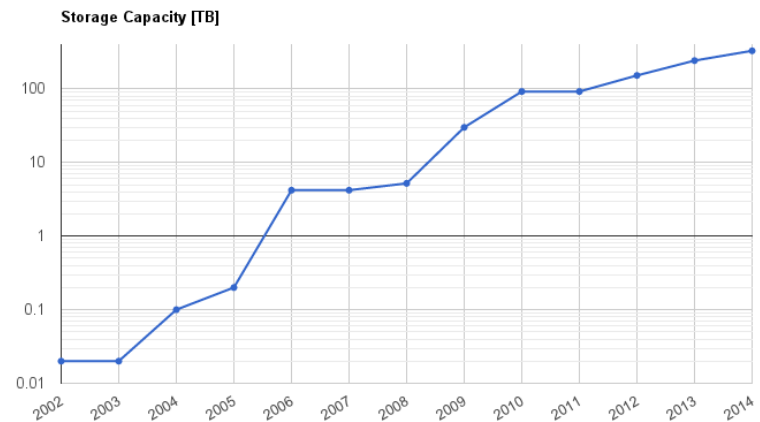
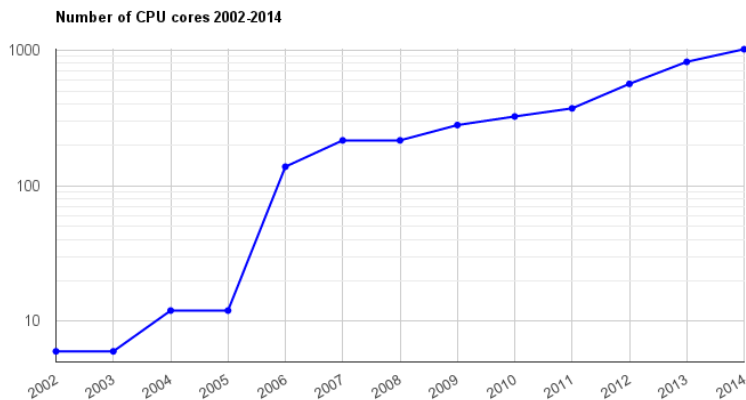


# Υπολογιστικές υπηρεσίες στο ΑΠΘ

- 2 υπολογιστικές συστοιχίες
  - διασυνδεδεμένες με το EGI μέσω Computing Grid
  - Περισσότερα από >1000 CPUs (Job slots)
  - Περίπου 3TB RAM
- Κεντρικές υπηρεσίες για το EGI
  - Catch-all CA (PKI)
  - Central VOMS (AAI – Authorization services)
  - Software repository ([repository.egi.eu](http://repository.egi.eu))

# Στατιστικά

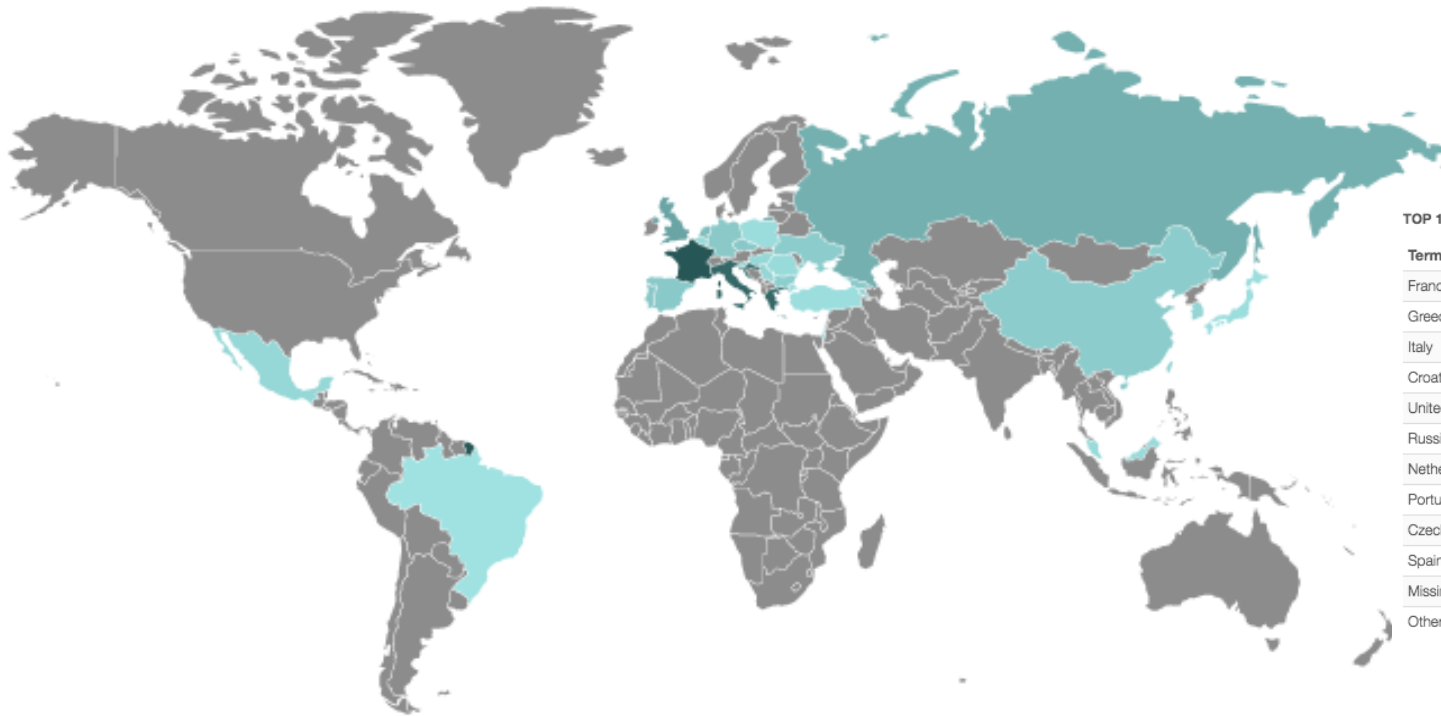
Μετρική	Μηνιαίως (περίπου)	2015
Number of jobs	50-70K	778,564
Utilized memory [TB]	60-70	784
Cpu Hours	200-300K	3,346,097
Nr. Of research groups	-	45
Nr. Of active users	-	121



# Utilization



# Origin of jobs

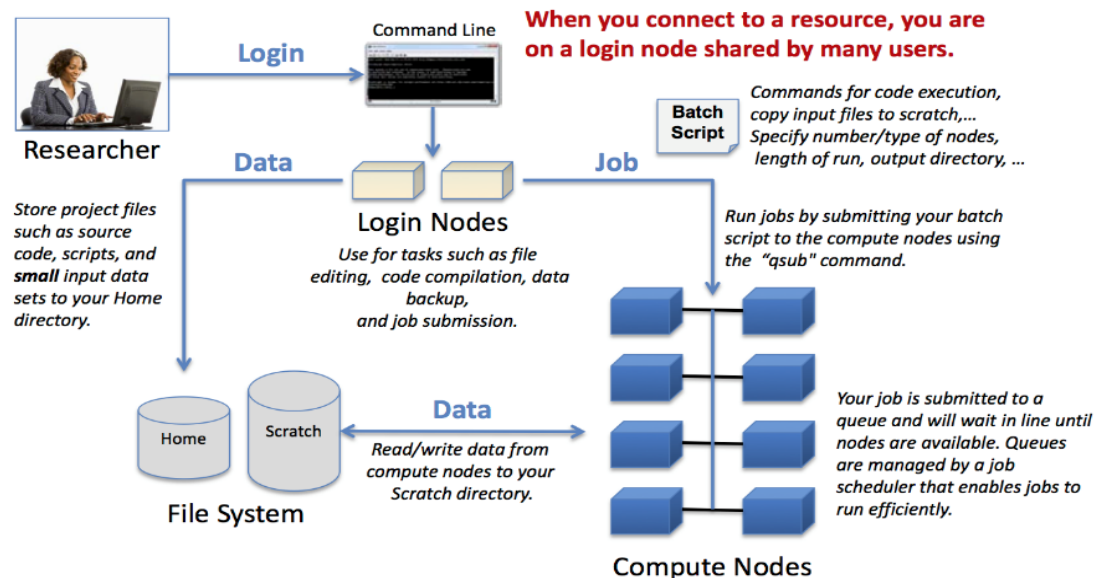


TOP 10 TERMS IN FIELD GEOIP.COUNTRY\_NAME

Term	Count
France	4909
Greece	3124
Italy	2972
Croatia	2882
United Kingdom	290
Russian Federation	156
Netherlands	72
Portugal	37
Czech Republic	33
Spain	32
Missing field	5282
Other values	268

# Υπηρεσία Compute (@Ιδρυματική συστοιχία)

- Χρήση Ιδρυματικής συστοιχίας:
  - Υπηρεσία για χρήστες (μέλη) του ΑΠΘ
  - Απλούστερες διαδικασίες και εργαλεία για νέους χρήστες
  - Μικρότερα queuing times (όχι πάντα αλλά συνήθως...)
  - Άμεση παραλαβή αποτελεσμάτων από εργασίες



# Εφαρμογές και Εργαλεία

- 10+ γλώσσες προγραμματισμού και διαδικαστικές γλώσσες
  - (π.χ. MATLAB, R, Mathematica, Python κ.α.)
- 20+ εργαλεία & βιβλιοθήκες
- 50 επιστημονικές εφαρμογές σε 7 ερευνητικούς τομείς:
  - Επιστήμες Γης
  - Ρευστοδυναμική
  - CAE/CAD
  - Engineering
  - Chemistry & Materials
  - Life Sciences
  - Physics

# Υπηρεσία Αποθήκευσης δεδομένων

- Συστοιχία αποθήκευσης (Gluster – Open Source)
  - 350TB ωφέλιμο storage
  - Διεπαφές Gluster (FUSE), NFS, Samba
  - Οριζόντιο scaling
  - Οικονομικό και αξιόπιστο
    - Οικονομία κλίμακας για όλο το Ίδρυμα
    - Προοπτική για παροχή SLA (αρχές 2017)



# Χρήστες υπηρεσιών Compute και Storage στο ΑΠΘ

- 45 ερευνητικές ομάδες (ενεργές κατά το 2015)
  - 100+ ερευνητές
- 9 καταγεγραμμένα use cases (από 5 τμήματα του ΑΠΘ)
  - 5 διδακτορικές διατριβές
  - 4 διπλωματικές εργασίες (3 στα πλαίσια ΠΜΣ)
  - 4 ερευνητικά έργα
  - 10+ δημοσιεύσεις και συμμετοχές σε επιστημονικά συνέδρια

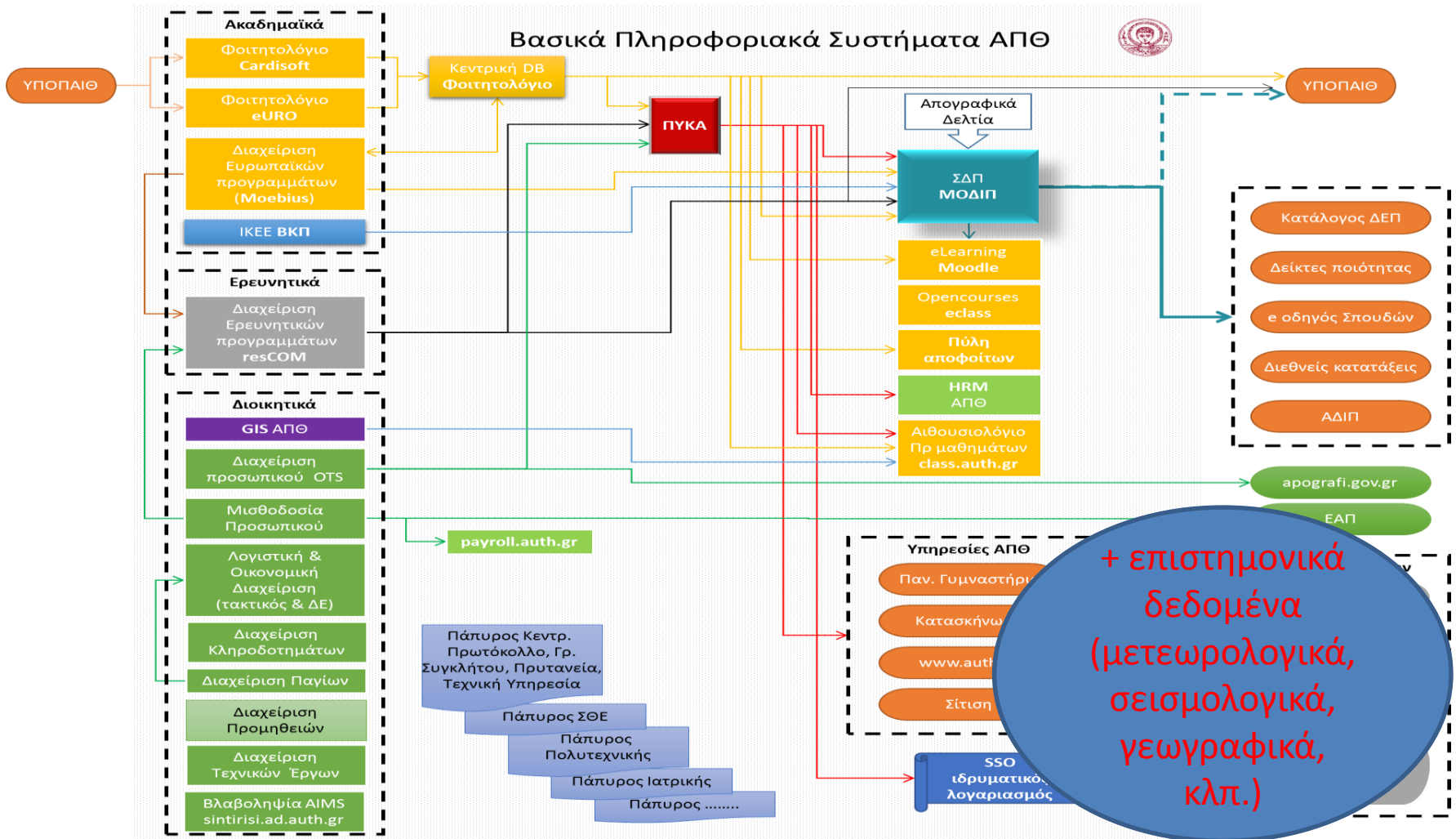
# Υπηρεσίες Big Data

- Με τον όρο Big Data υποκρύπτουμε συνήθως εργαλεία που αναπτύχθηκαν στο χώρο του IT για εσωτερικές ανάγκες μεγάλων οργανισμών. Παρόλα αυτά υπάρχουν εφαρμογές και στην έρευνα
- Δύο συστοιχίες
  - Hadoop
    - HDFS, YARN, Tez κ.α.
  - Elastic
    - Apache Solr
- Προοπτική για δυναμική παροχή Big Data εργαλείων σε χρήστες του ΑΠΘ (π.χ. Spark, Elastic, Hama, Hive κ.α.)

# Open Data στο ΑΠΘ

- Το ΑΠΘ είναι ήδη πάροχος ανοιχτών δεδομένων
- Πλεονεκτήματα από το άνοιγμα των δεδομένων:
  - Διαφάνεια
  - Βελτίωση της αποδοτικότητας και της αποτελεσματικότητας των υπηρεσιών
  - Ενίσχυση διαδικασιών αξιολόγησης (εσωτερικής και εξωτερικής)
  - Ανακάλυψη, συσσώρευση και αξιοποίηση νέας γνώσης από το συνδυασμό διαφορετικών πηγών σε μεγάλους όγκους δεδομένων

# Τα ανοιχτά δεδομένα του ΑΠΘ



# Περιεχόμενο ανοιχτών δεδομένων

- Τα ανοικτά δεδομένα του ΑΠΘ περιλαμβάνουν πληροφορίες για Διδάσκοντες, Πρόσωπα, Προγράμματα Σπουδών, Μαθήματα και πληροφορίες σχετικά με Ακαδημαϊκές και Διοικητικές Μονάδες του πανεπιστημίου
- Διαθέσιμα μέσω του [data.gov.gr](http://data.gov.gr)
- Αναλυτικές περιγραφές και παραδείγματα χρήσης:
  - <http://www.data.gov.gr/organization/aristotle-university-of-thessaloniki>

# Open Data για ερευνητικές ανάγκες

- Αξιοποίηση των δεκάδων TB ερευνητικών δεδομένων
- Προσθήκη http/ftp front-ends
  - Data curation από ερευνητικές ομάδες
    - Endorsement οντολογιών από schema.org (π.χ. DataSet)
  - Προσθήκη λειτουργικότητας στο υφιστάμενο API
    - <https://ws-ext.it.auth.gr/>
  - Χρήση PID service (registry)
- Οφέλη:
  - Data references
  - Data citations

# Open Science

- Προσβασιμότητα σε ερευνητικές δημοσιεύσεις (open access)
  - Προσβασιμότητα σε δεδομένα
  - Προσβασιμότητα σε αλγορίθμους (open source κώδικας)
- ➔ Δυνατότητα αναπαραγωγής και ανασύνθεσης ερευνητικών δεδομένων & αποτελεμάτων

# Σχεδιασμός



Εργαλεία επεξεργασίας,  
ανάλυσης και  
οπτικοποίησης

Δυναμική παροχή  
υπολογιστικών πόρων



Αποθετήριο  
ανοιχτού κώδικα

Αποθηκευτικές  
συστοιχίες/  
υπηρεσίες



Open Data API





# Χρήσιμα links

- Υποστηριζόμενες εφαρμογές
  - <https://it.auth.gr/el/software/sciComputingSoftware>
- Αίτηση για υποστήριξη εφαρμογών
  - [https://it.auth.gr/el/eForms/complimentary\\_scientific\\_software](https://it.auth.gr/el/eForms/complimentary_scientific_software)
- Use cases
  - <https://it.auth.gr/el/sciComp/research>
- Επικοινωνία
  - [scientific-computing@it.auth.gr](mailto:scientific-computing@it.auth.gr)