



# DATA LIFE AND OPEN DATA PARADIGMS

5<sup>TH</sup> APRIL 2018

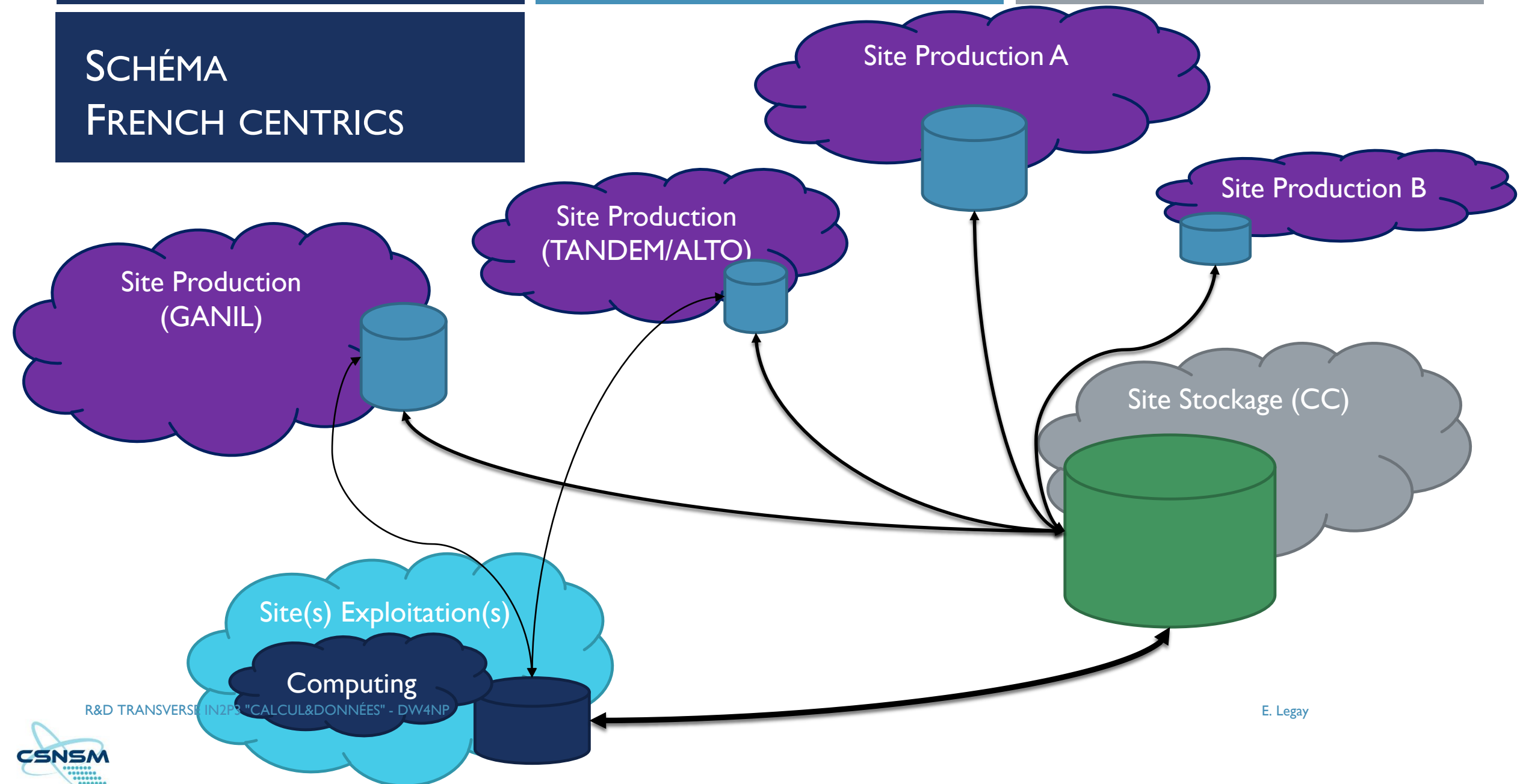
# WHAT MEANS OPEN DATA ?

- « Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike. »
  - Availability and Access : data have to be available at no cost and in convenient form,
  - Re-use and Redistribution,
  - Universal Participation : restrictions of use for certain purposes are not allowed.
- Why are we impact by the « Open data » paradigm ?
  - Funded by States / Institute / ...
  - May 2016 : European Council calls to **immediate** open access to scientific papers and datas ... in **2020**.
  - Data Management Plan (DMP) is mandatory to submit H2020 project
  - Hot point (at least in France for now) ...
    - IN2P3 and CEA makes pressure on GANIL to have DMP before end of year
    - IN2P3 funds technical project to improve data management in nuclear physics

# IN2P2 - DATA LIFECYCLE FOR NUCLEAR PHYSICS

- Way to manage data produced during last 20 years is ....At least awful
  - Tapes lost somewhere under dust
  - Logbook ... somewhere but not under the same dust ... if not trashed !
  - Physicist remembering how to analyse the data ... under paperwork for new experiment

# SCHÉMA FRENCH CENTRICS



# IN2P3 – A NEW 2 STEPS PROJECT

- Step 1 – Sending data to common data center
  - Needs of DMP
  - **Web proxy to retrieve data**
    - Meta data to identify data sets
    - Rights procedures (based on DMP) to give access to datas
    - Centralized authentication multi sites, multi collaborations, multi ...
- Step 2 – Analyze data center
  - How to access to data ?
  - How to provide right tools for analyze ?
    - Based on batch mode farms ?[[[[
    - Based on IaaS computing cloud ?
    - Both ?