

# HOW TO BETTER EXPLOITING THE DATA MANAGEMENT AND DATA ENGINEERING EXPERTISE OF RESEARCH INFRASTRUCTURES FOR THE DEVELOPMENT OF THE EUROPEAN OPEN SCIENCE CLOUD TO FOSTER DATA SCIENCE AS WELL AS INCREASED DATA SHARING AND USE BEYOND THE INDIVIDUAL RIs OR SCIENTIFIC DOMAINS?

## WHITE PAPER MAIN MESSAGE # 5

### **GIORGIO ROSSI**

Università degli Studi di Milano, Dipartimento di Fisica

NFFA-Europe Coordinator

EOSC GB member

*HIGH LEVEL EXPERT GROUP REPORT ON RI - LTS (2020):  
SUPPORTING THE TRANSFORMATIVE IMPACT OF RIs ON EUROPEAN RESEARCH*

- The RIs are at the crossroads of data-producing facilities that need to **constantly upgrade their instrumentation** to stay at the forefront of research, including the e-infrastructure that goes with it and the **EOSC-like services** that will be made accessible also under the **umbrella services of EOSC**. As the EOSC services may generate a traffic of data science work exceeding the capacity needs of each RI for its own science programme, criticalities can be expected on both sides of RIs and EOSC.
- Many RIs, EIROforum Members, ESFRI, ERICs or other large national RIs, are in leading position as building blocks of the development of the EOSC and as participants of H2020-Clusters. ELIXIR, EPOS ERIC, ESRF, CEESDA ERIC and others, including H2020 IAs, provide ***state-of-the-art references of the most advanced data management solutions*** for fulfilling the FAIR principles and feeding quality controlled datasets into the future EOSC.
- The implementation process of FAIR data and related services and the goal of interoperability of data across domains, provides a stimulus towards the full integration of the RIs into an ***overall interoperable research system***, offering data and coordinated access to the production of new datasets in response to both disciplinary research needs and multi or interdisciplinary demands arising from complex problems as the great challenges or societal challenges, ***like the current pandemics***.

- RIs are today often at the forefront of data formatting, curation, archiving and access, but less so on the **interoperability aspects**, starting from the elaboration of high-quality metadata. The effort towards interoperability has structural and financial implications that must be addressed with the RIs, in a concerted and well supported manner.
- **Sustainability of the EOSC and of the RIs are interlinked. Very likely the real way to an effective EOSC implies that all new datasets must be FAIR from the acquisition stage (FAIR-by-design) as the FAIR retrofitting of existing datasets will be limited to essential time-series or to observational data that intrinsically cannot be acquired again. 'FAIRification' of existing data sets will imply criteria of choice, or a system enabling the FAIRification on-demand, provided a metadata open access environment is established at the core of EOSC.**
- Acquiring **FAIR data by design** implies large instrumental investment (hardware and software) as the complementary data to create metadata must be **acquired in automatic manner** to reduce the extra work required to the researchers to a sustainable level.

An example is the current effort, stimulated by EOSC and implemented by EMBL-EBI, for creating a COVID-19 Platform:



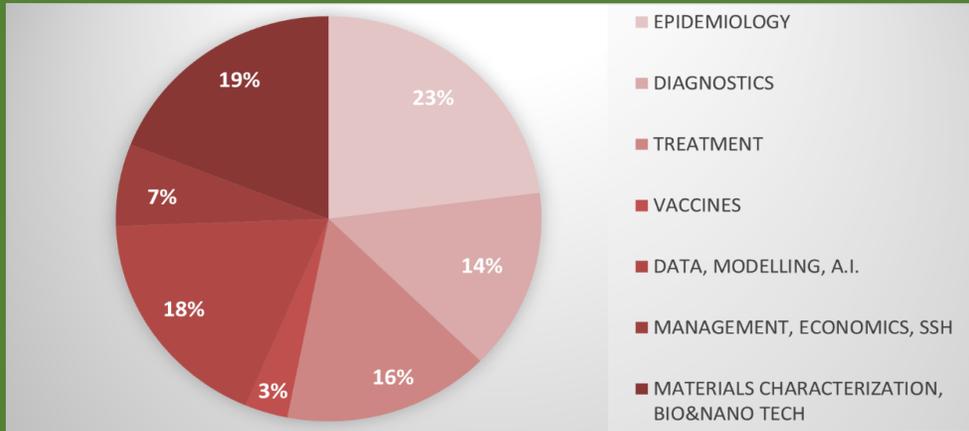
[About](#) [Data Hubs](#) [Related resources](#) [Our partners](#) [Submit data](#)

[Sequences](#) [Expression](#) [Proteins](#) [Structures](#) [Literature](#)

Accelerating research through data sharing

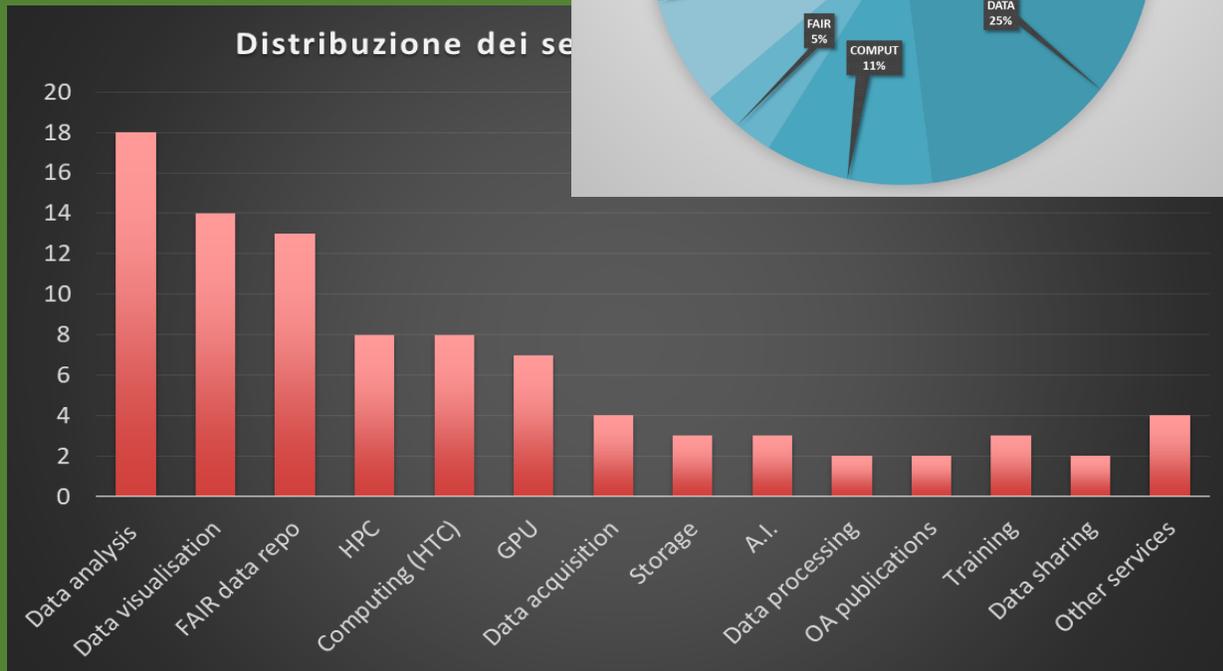
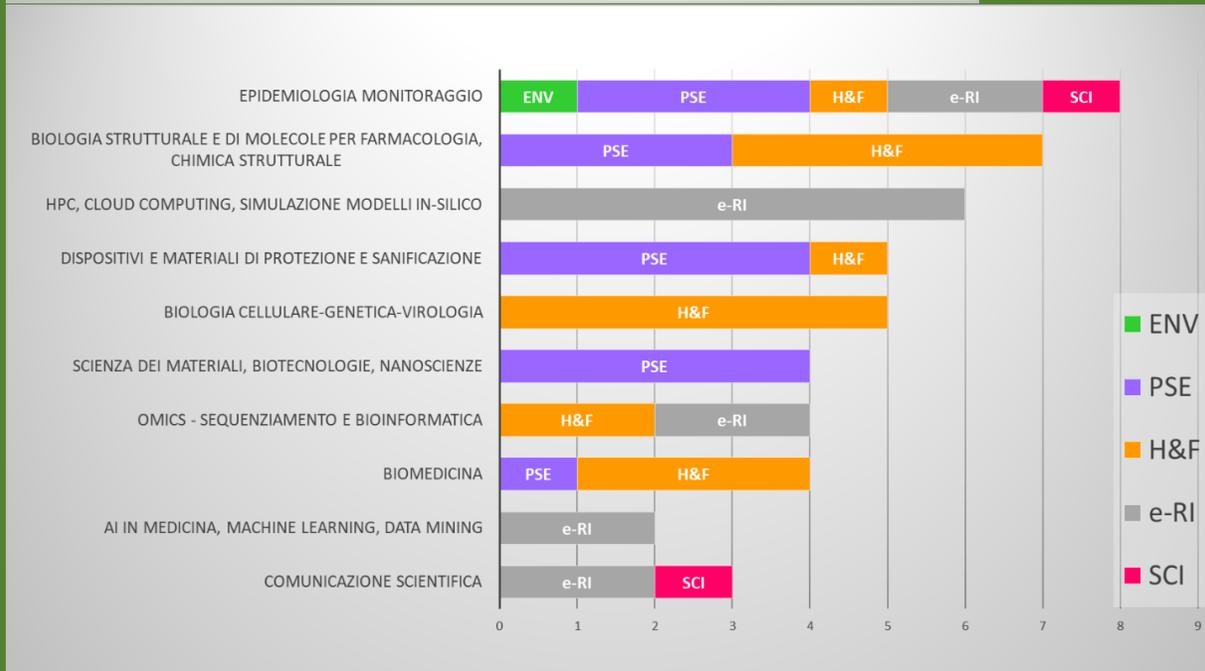
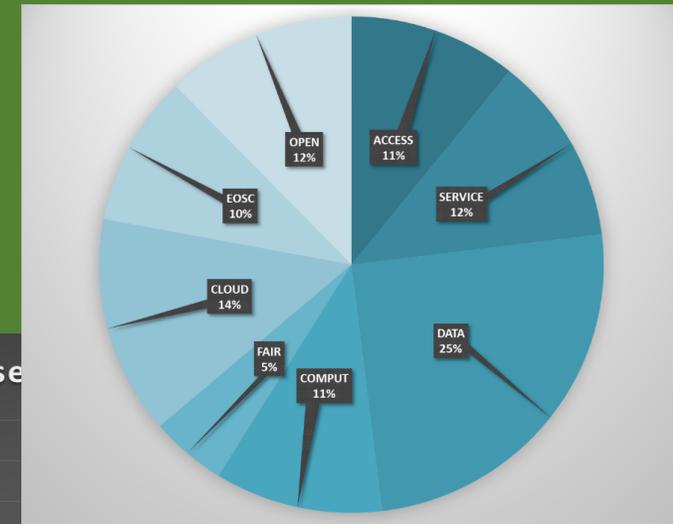


- At national level the effort of RIs and Data Infrastructures to face the urgent needs of research on SARS-CoV-2 and COVID-19 is also producing a real-time response (ICDI-ESFRI survey in Italy, May 2020):



access opportunities to support research projects

services offered by RIs



- **INTEROPERABLE SYSTEM OF RIs must be implemented to respond to complex issues as EMERGENCIES and MISSIONS**
- **DATA AND NEW-DATA puts both EOSC AND RIs AT WORK**
- **THE WAY RIs AND EOSC ARE REACTING to to COVID-19 IS INSPIRING OF THE INTEROPERABLE RESEARCH SYSTEM**
- **RIs of DIFFERENT DOMAINS ARE SUPPORTING RESEARCH FOR SARS-CoV-2 and COVID-19**
- **REMOTELY ACCESSIBLE RESOURCES, INTELLIGENT-INTERACTIVE REMOTE ACCES, FAIR-by-design NEW DATASETS**
- **INCREASE MAN-POWER at RIs, as a RECOVERY MEASURE FOR EU RESEARCH COMPETITIVENESS to perform REMOTE ACCESS**
- **COVID-19 IS AN EXAMPLE OF URGENT NEED OF NEW KNOWLDEGE AND NEW RESEARCH**
- **There is EVIDENCE of HIGHLY INTEGRATED/INTERACTING COMMUNITIES as well as of the NEED OF DATA that are today NOT STANDARDIZED : epidemiology data, human phenotype data, clinical data. THERE ARE BARRIERS - legal, cultural, economic...- which cause the CHAIN OF NEW KNOWLEDGE to be somewhat interrupted.**
- **USERS are CURISITY DRIVEN which guarantees excellent science, BUT USAGE OF RIs SHOULD ALSO be MISSION-ready though interoperability of both RIs and Data**