# Some [unstructured] remarks from the Scoping Paper

Davide Salomoni (davide@infn.it)

LODES 2018

Frascati, Sep 6, 2018

The following is a rather unstructured dump of remarks related to the [current draft of the] Scoping Paper. Details in the slides might be missing for either the sake of time or just because of my ignorance.

I come from a physics/technological background, so please forgive me if my notes will be mostly oriented to technology.

# Synergies

- Focus on the **synergies & discovery of what is available**.
- This is also because we have a clear lack of common solutions that allow interoperable, linked open data across multi-disciplinary communities.
- This is, I believe, actually one of the big issues facing any EOSC-related endeavor.

# Data lifecycle

- Consider the **full data lifecycle**. This is a complicated issue, mostly handled manually if at all today.
- It relates directly to **SLA**, for example because of **dynamic (time-based) authorization policies**.

# Replication & preservation

- **Open data replication and long-term data preservation** is another area that deserves attention. It has direct impact on how is data stored, preserved and accessed.
- Caches (content delivery networks for data) might be useful here. But we do not have an off-the-shelf, open, scalable solution for this problem

# Building bridges

- While I think we need to move beyond general "explanations" about open data even in reports aimed to policy makers, we also need to strive to **move away from purely "technocratic" (or "bleeding edge") solutions**.
- This, in the sense that even if we found workable solutions for some issues, we need to present and test them ASAP:
  1. in an agile way, and
  2. involving non IT-specialists from the very start. This will also help in simplifying the provisioning of these solutions by resource providers. But users are key, not only for requirement collection, but also for early testinga and co-development (maybe through smallish hybrid, international testbeds).

# Data orchestration

- Collaborate with **other projects or infrastructures that work on (in some cases, directly provide) data orchestration solutions**. This is a key point, and it is not trivial at all from a technological point of view.

# Data reproducibility

- When considering data quality, it is key to also focus on **data reproducibility**. This is particularly important when using open data to derive other open data, and is a gigantic issue in some scientific domains.

# The real value: people

- How can career progressions be encouraged for those who work on these topics? But, more in general, how can **research on open data topics be supported with a model that goes beyond a pure project-based paradigm?**

# Thank you

“Stat rosa pristina nomine, nomina nuda tenemus”