

Roles of e-Infra in a changing world

Reflections after 25 years of e-Infrastructure building

Peter Wittenburg, Paolo Budroni
FDO Forum, TU Wien, ASSAI

Looking back on 25 years of e-Infrastructure building

- Many funds available at EU and MS level to explore and improve the data landscape extensively and to get many people activated (DOBES, CLARIN-ESFRI, EUDAT, EOSC, NFDI, FDO One, etc.)
- Some initiatives co-started such as RDA, FDO Forum
- Some documents produced (Riding the Wave, FAIR Principles, etc)
- **Still not fully clear how to optimally organise the data landscape**
 - Amount of available funds will decrease
 - Pressure to come to decisions will increase
 - Bottom-up concept will not work – too many smart people out there

New key-words added: *sovereignty and self-sustained*

- GEANT and PRACE/EuroHPS found their role – mapping geographical with technological structures and funding aspects
 - **Data is different**
 - Connecting data in trusted environments is much more complex than connecting computers
 - There are rich, complex and multifaceted landscapes of institutions and communities with great ideas and a huge number of tools/solutions (national differences) which will NOT change overnight
 - Data has become integral and fundamental part of trusted scientific workflows
 - **Role of data in science is dramatically changing (large + small AI models)**
- **Need for sovereignty, transparency/traceability, accountability and persistence**

Need for „central“ and trusted e-Infra services in data domain?

- Let's first understand some achievements

Major achievements of ESFRIs (NFDIs)

- organised the community at service provider level and unlocked huge amounts of data from „dark space“
- built their „trusted dataspace“ i.e. agreed on aggregation technologies for metadata harvesting, *on sets of policies, etc.*
- developed portals and tools (technologie will come and go)
- **created social knowledge and a surrounding trust framework**

New landscapes of challenges

Today some experts are using infrastructure services

... however AI is changing the game: increasing number of small and large models to come...

Let's take EPOS as an excellent example:

- after the integration of metadata - now the integration of data as follow-up step
- *smaller* models emerge – no particular challenge within communities
- smaller models across communities are hardly possible
- is Europe prepared to use its great potential for large domain models??

General Problems:

- Need for efficient exchange of data vs. controlled access
- Who owns the services on large (and small) data models

Requirements for an efficient domain supporting sovereignty

→ We need a domain of FDOs and easy to deploy „dataspace technology“

But, why FDOs ?

→ Persistent references, bundling/encapsulating all information about a data items incl. the rights (ODRL assertions etc.)

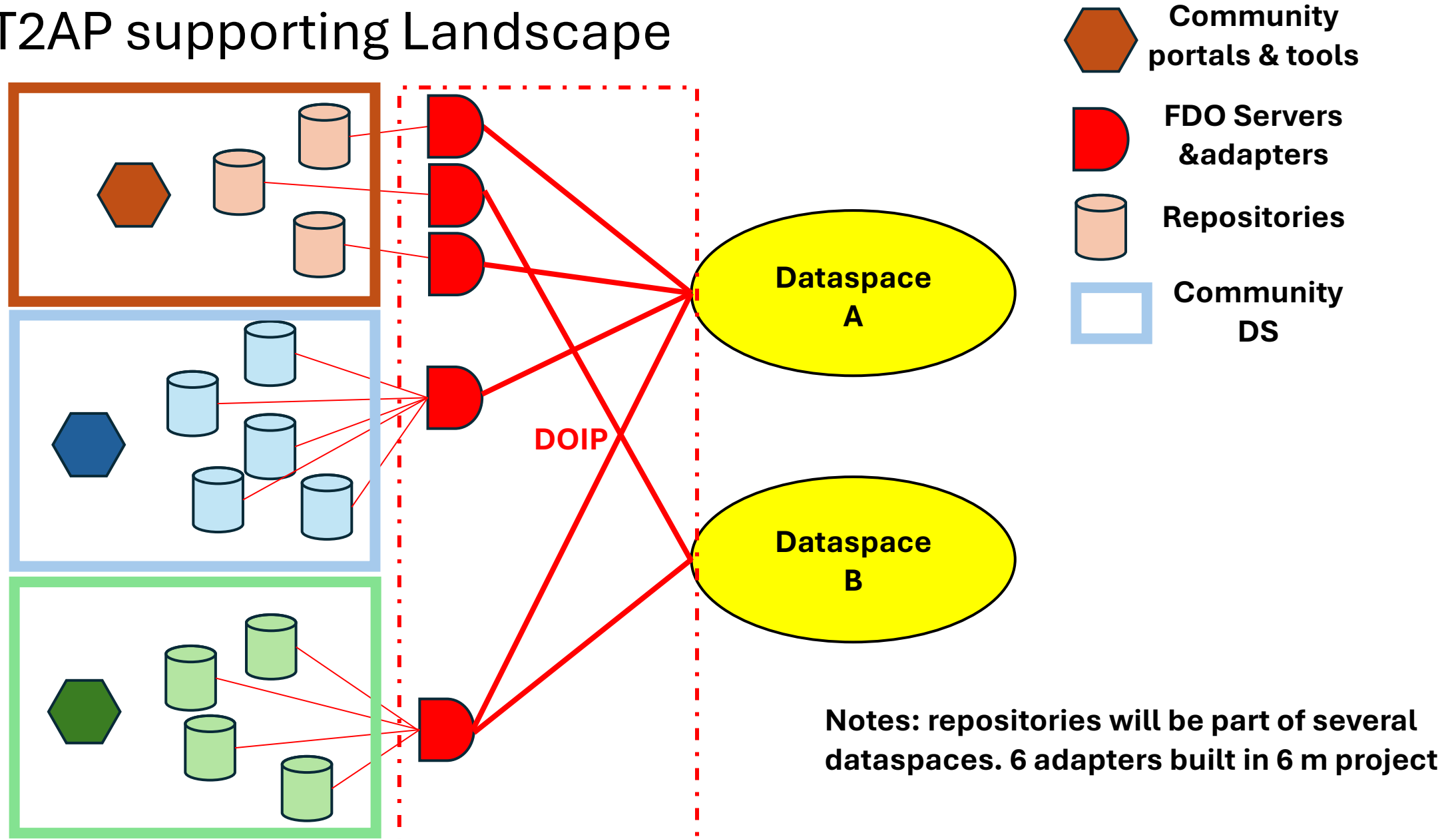
→ self-standing autonomous units, can even become autonomous agents, build the basis for trust

→ *FDOs encapsulate rights, but do not interpret them !!!*

And why Dataspace Technology (DST) ?

- **It's about interpreting rights and turning assertions to actions**
- Always needs to consider national laws and ethics
- DST has three layers: trust building – rights/claims negotiation – exchange
- FDO informs DS on all layers
- *Note: several DS technologies are existing or rapidly emerging*

The ST2AP supporting Landscape



Possible Tasks of e-Infrastructures

Providing basic elements of such a trusted infrastructure such as

- Easy to use, flexible and affordable PID system (may be competitive)
- Maintaining a couple of registries for FDO work
- Developing and enhancing trust frameworks (including non-technical issues like training and skills development)
- Support for easy to deploy DS technologies
- **Workspaces to aggregate training data and cycles to built models**

.Setting up global trust enhancing governance structures where necessary