

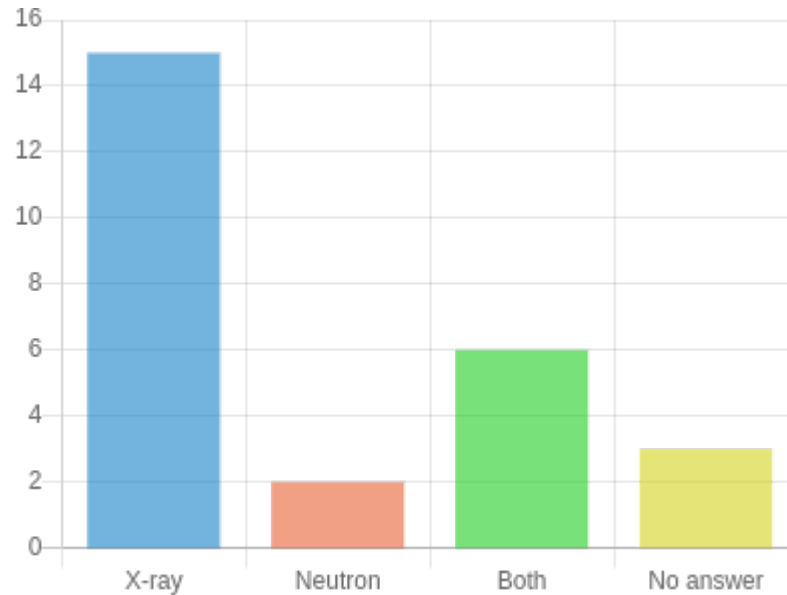


Leveraging open data from PaN facilities for machine learning – day 2

09:00	Introduction and objectives day 2			<i>Sophie Servan</i> 09:00 - 09:15
	Metadata requirements for single crystal raw diffraction data re-use. Experience of IUCrData - Raw Data Letters. <i>Loes Kroon-Batenburg</i>			
10:00	Parallel working session: Working group 1	Parallel working session: Working group 2	Parallel working session: Working group 3	
11:00	<i>Principal/1-57 - Salle A1.1.57/Orion, Batiment Principal 09:45 - 11:45</i>	<i>Principal/1-22 - Salle A2.1.22/Pyxis, Batiment Principal 09:45 - 11:45</i>	<i>Principal/1-48 - Salle A1.1.48/Virgo, Batiment Principal 09:45 - 11:45</i>	
12:00	Report to the group			<i>Majid OUNSY et al.</i> 11:45 - 12:15
	Workshop conclusion			<i>Sophie Servan</i> 12:15 - 12:30

Survey results – part I

Which type of facilities are you using or are you interested in using?

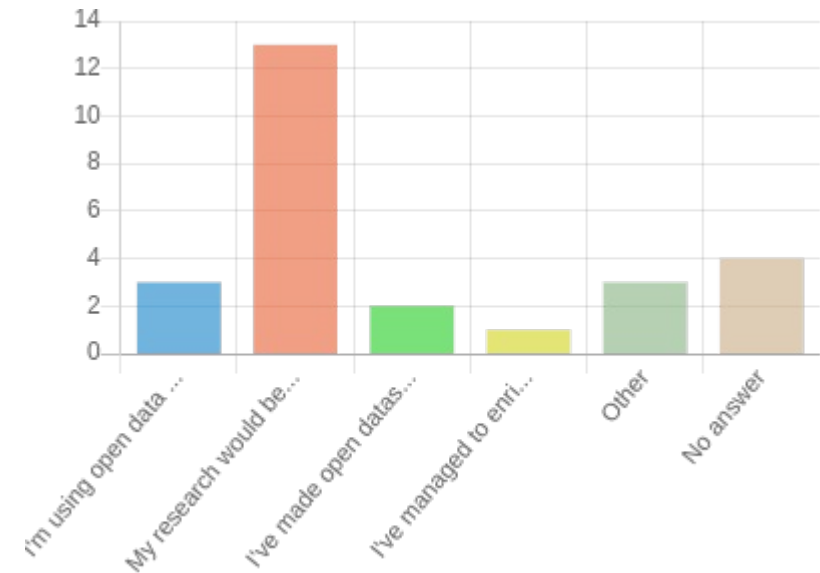


Top techniques:

- X-ray powder diffraction (3)
- Crystallography (3)
- Small-angle scattering (3)
- Reflectometry (3)
- X-ray absorption (3)
- Tomography (2)
- Total scattering (2)

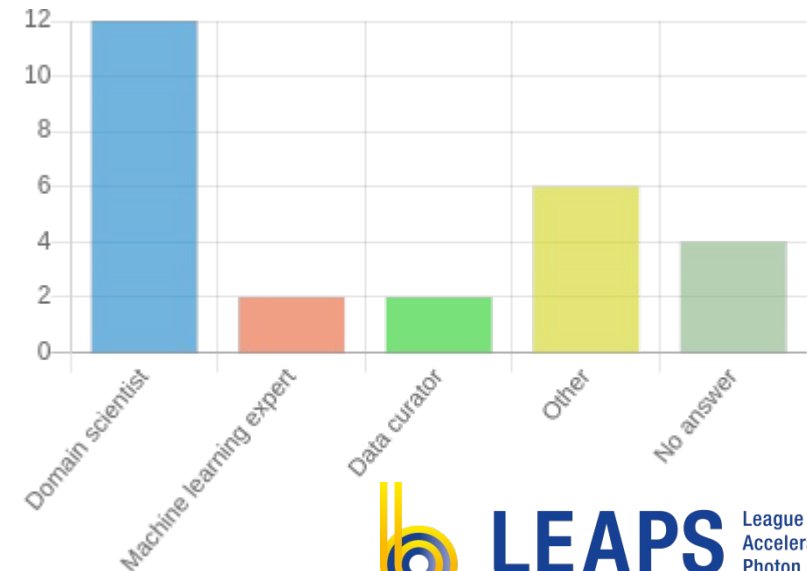
24 responses

Which point of view are you mostly interested in?



* My research would benefit from more and better curated open data

What best describes your position?



Survey results – part II

Among the following thematic areas, please select the ones you find interesting, ranking your preferences.

- **Harnessing open data for machine learning: curated repositories, pre-processing, reproducibility (46)**
- **Publishing qualitative datasets: metadata, citation, licence, quality-checks (41)**
- **Collaboration between data producers and consumers: challenges and opportunities (34)**

Other proposals:

- Software: a good practice in publishing software leveraging ML for photon science data, workflows, pipelines. (1)
- Repurposing/retraining existing models for one's goals. (1)

24 responses

Working groups

- **#1: Harnessing open data for machine learning:** curated repositories, pre-processing, reproducibility
Chair: **Paul**
Room: **ORION // breakout room 1**
- **#2: Collaboration between data producers and consumers:** challenges and opportunities
Chair: **Markus**
Room: **PYXIS // breakout room 2**
- **#3: Publishing qualitative datasets:** metadata, citation, licence, quality-checks
Chair: **Majid**
Room: **VIRGO // breakout room 3**
- Links to collaborative notes in Indico!