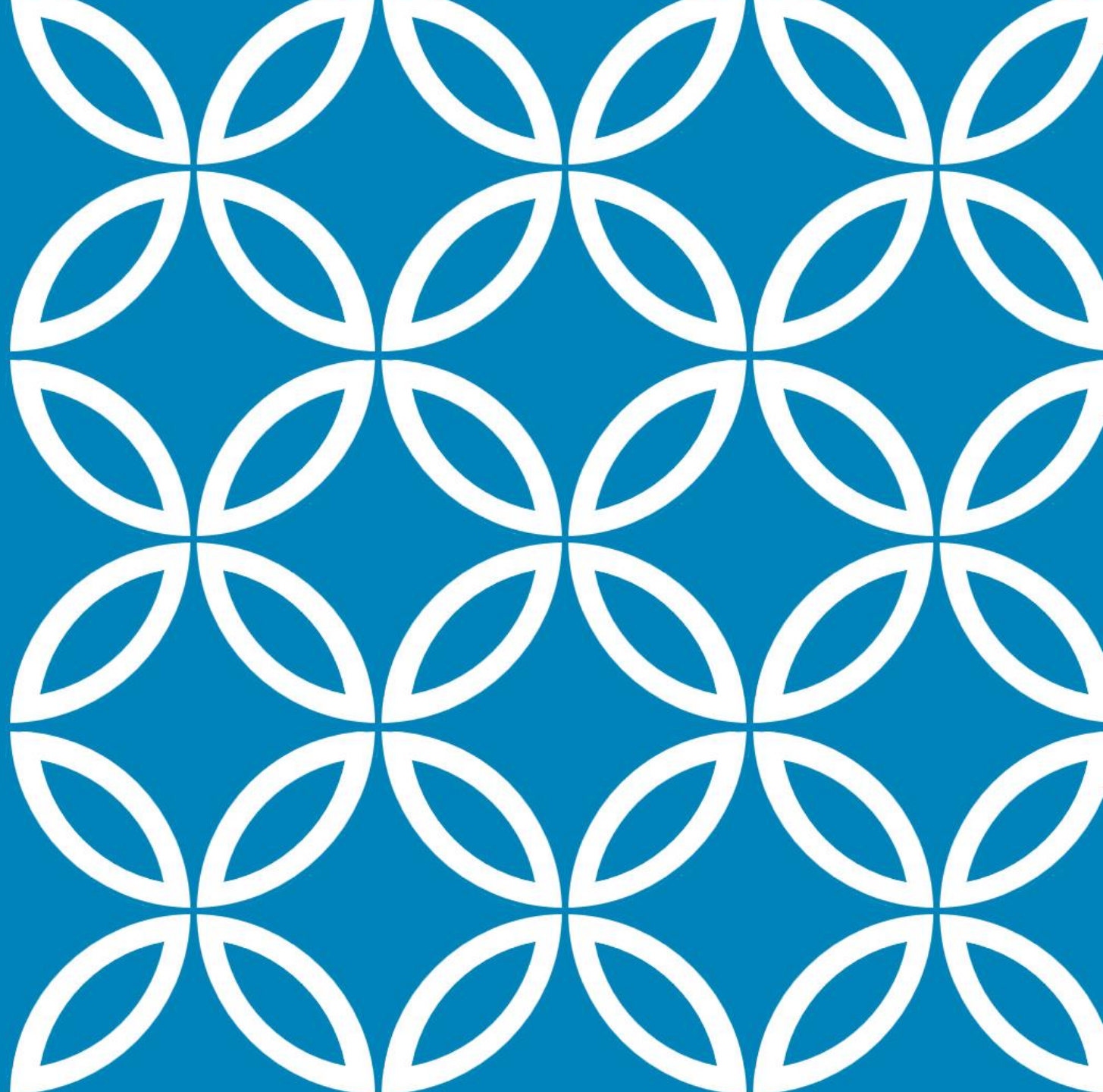


# CURRENT PRACTICES IN (HOME LAB) RAW DATA ARCHIVAL

---

Responses to a survey published  
at:

<https://www.iucr.org/news/newsletter/volume-28/number-1/raw-data-availability-the-small-molecule-crystallography-perspective>





# SURVEY DETAILS

Survey run in 2019 – widely distributed through various networks:

- IUCr website
- Regional Crystallographic Associations
- Vendor user lists
- Personal word-of-mouth

## Responses

- 189 useable responses (one was incorrectly recorded)

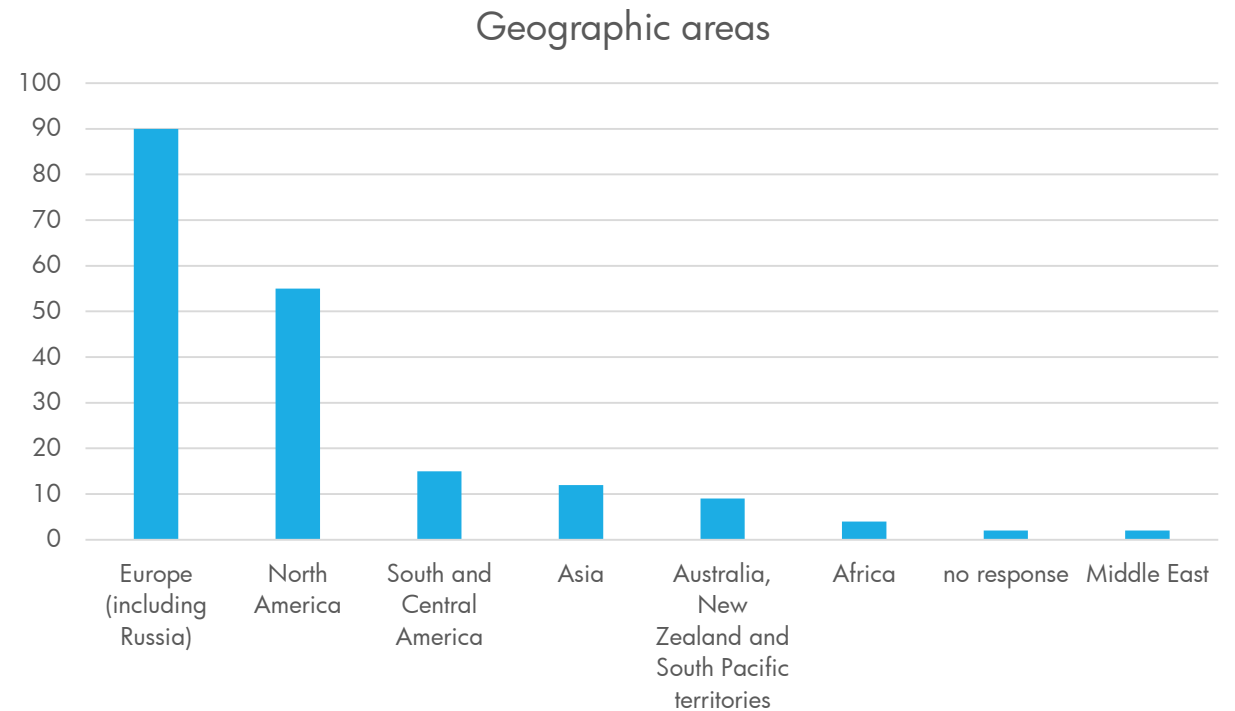
Thank you to everyone who responded and encouraged others to respond!

# DEMOGRAPHICS – REGIONAL AFFILIATION



At least one response from each region.

Three-quarters of responses came from Europe + North America.

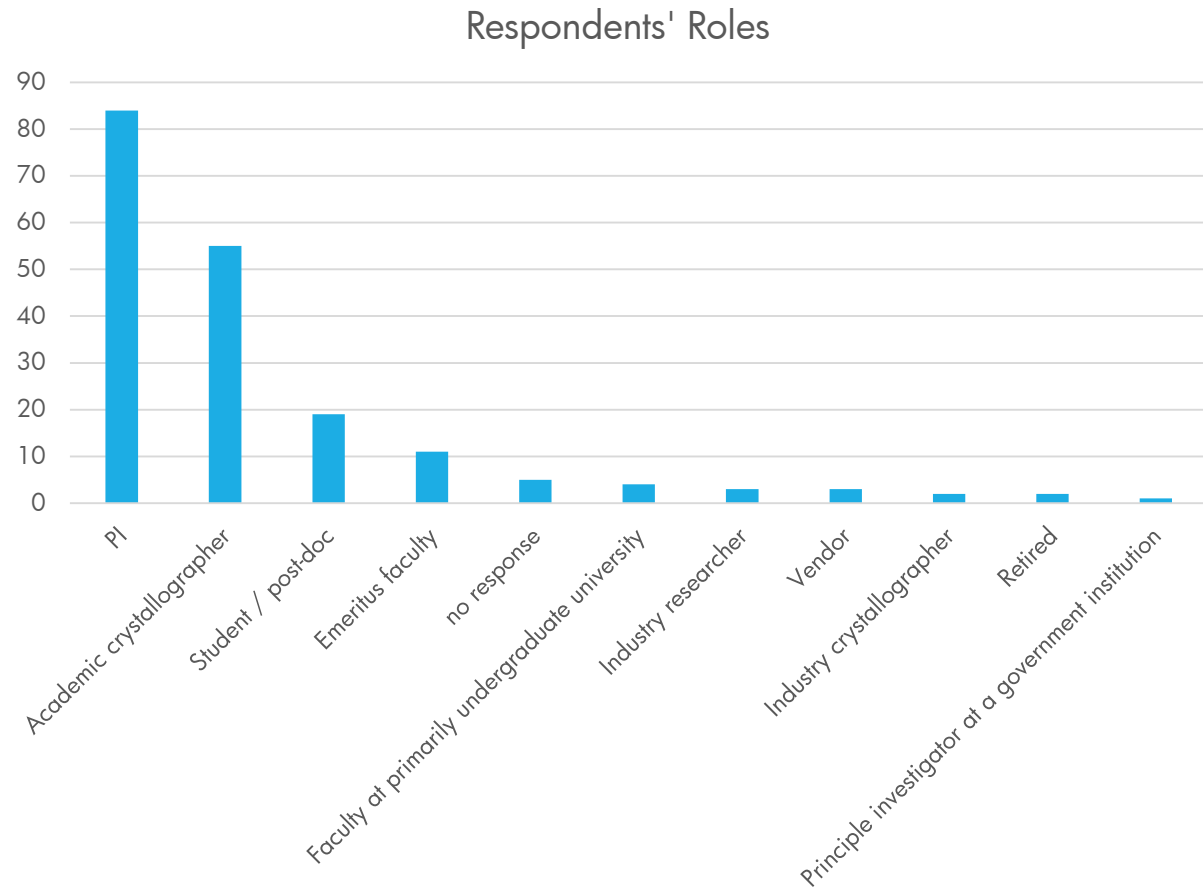


# DEMOGRAPHICS – RESEARCH SECTOR



Majority response from  
Academia

Few from government/industry



# REASONS FOR ARCHIVAL/MANAGEMENT



## Validation:

- Quality (135)
- Scientific claims (111)

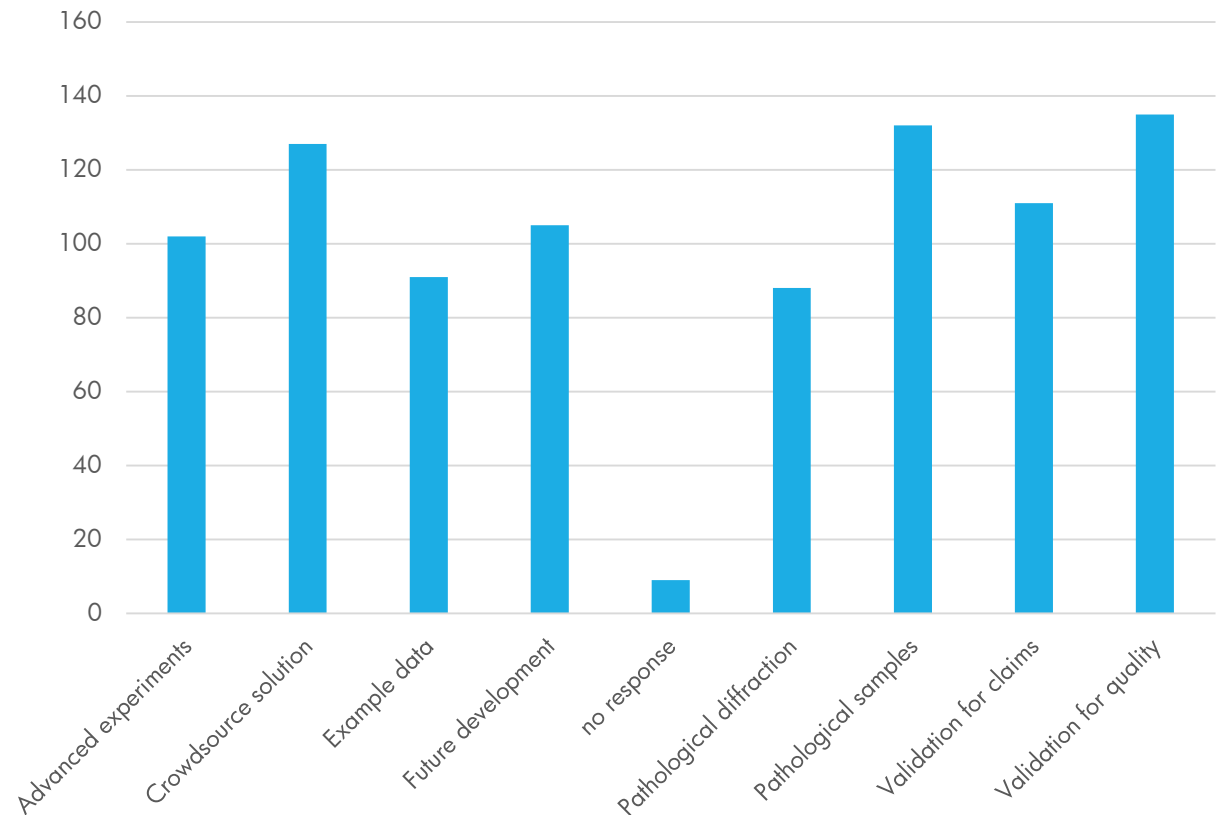
## Understanding:

- Providing example data (91)
- Pathological samples (132)
- Crowdsolve solutions (127)

## Development

- Future developments (105)
- Advanced experiments (105)

Worthy reasons for raw data archival





# USES OF RAW DATA

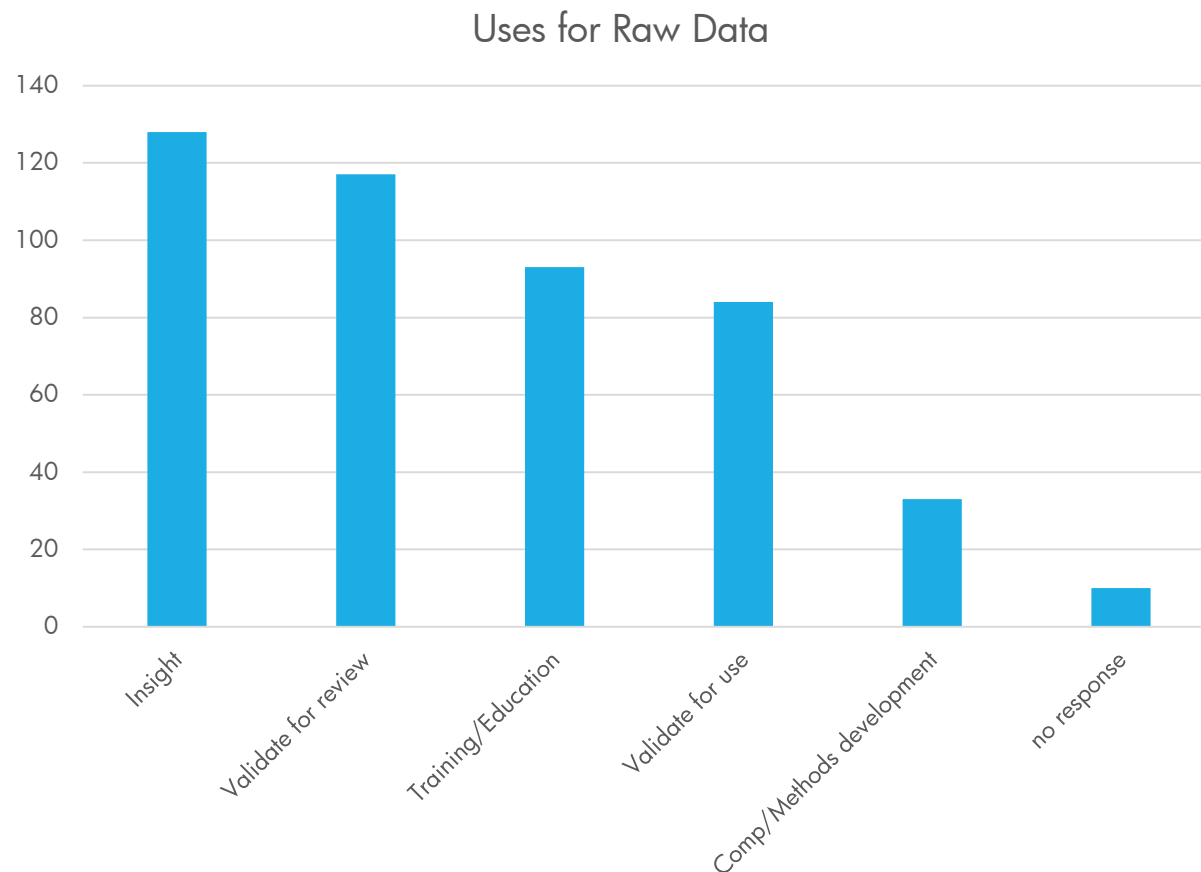
Majority agreed:

- Scientific insight (128)
- Validation during review (117)

Less than half agreed:

- Training/Education (93)
- Validating data for (re)use (84)
- Computational work/Method development (33)

A few individual write-in comments from those who already have enough data of their own, trust the data contained in the CIF, or see no need for accessing others' raw data (6 in total).





# CONCLUSIONS?

Responses are skewed:

- Europe/North America
- Academia vs. Industry/Government

General understanding of the utility of widely available raw data

- Validation
- Education
- Development

Today's discussions will focus on use cases for publicly available raw data

Tomorrow will focus on current practices and future directions

# TODAY'S AGENDA



<b>Day 1: Wednesday 11 August</b>		
<b>Session 1: Use cases highlighting the value of raw data management/availability in SMX</b>		
<b>14:00-14:10</b> (08:00-08:10)	<b>Amy Sarjeant</b>	Workshop introduction and results of the CommDat survey on data management and availability
<b>14:10-14:35</b> (08:10-08:35)	<b>Michal Dušek</b> (Prague)	The value of the availability of raw data for (re)determining modulated/incommensurate structures
<b>14:35-15:00</b> (08:35-09:00)	<b>Kamil Dziubek</b> (Poznan)	Using raw data to improve the interpretation of high-pressure diffraction experiments
<b>15:00-15:25</b> (09:00-09:25)	<b>Amber Thompson</b> (Oxford)	Making a Silk Purse out of a Sow's Ear: Publishing Difficult Data
<b>15:25-15:45</b> (09:25-09:45)	<b>Coffee</b>	
<b>15:45-16:10</b> (09:45-10:10)	<b>Simon Grabowsky</b> (Bern) <b>Krzysztof Wozniak</b> (Warsaw)	Inclusion of raw data to support validation and publication in quantum crystallography
<b>16:10-16:35</b> (10:10-10:35)	<b>Jim Britten</b> (McMaster)	Use of raw data for diffraction space visualization
<b>16:35-17:00</b> (10:35-11:00)	<b>Joe Ferrara</b> (Rigaku)	Instrument vendor perspective – what could be done?
<b>17:00</b> (11:00)	<b>Close of Day 1</b>	



# TODAY'S AGENDA



<b>Day 2: Thursday 12 August</b>		
<b>Session 2: What would a 'standard' look like? – best practice</b>		
<b>14:00-14:10</b> (08:00-08:10)	<b>Simon Coles</b> (Southampton)	Current practices in raw data archival (home laboratories)
<b>14:10-14:35</b> (08:10-08:35)	<b>Graeme Winter</b> (Diamond)	Future outlook for a curated raw data archive
<b>14:35-15:00</b> (08:35-09:00)	<b>Loes Kroon-Batenburg</b> (Utrecht)	How to make structures better in the future and the use of published raw data in crystallographic software development
<b>15:00-15:20</b> (09:00-09:20)	<b>Coffee</b>	
<b>15:20-15:45</b> (09:20-09:45)	<b>Teodor Ivănoaica</b> (Extreme Light Infrastructure)	Data Policy at Large Research Infrastructure. Data Retention challenge
<b>15:45-16:10</b> (09:45-10:10)	<b>Natalie Johnson</b> (CCDC)	The value of integration of raw data with publishing procedures and results databases
<b>16:35</b> (10:35)	<b>Panel discussion</b>	<b>16:35</b> (10:35)
<b>17:00</b> (11:00)	<b>Close of Day 1</b>	