

Improve your DMP with expert guidance

CESSDA Webinar

10.5281/zenodo.10118731

Lisa Hirsch, AUSSDA

Anne Sofie Fink, DEIC

Hannah Mihai, DEIC

14 Nov 2023

 cessda.eu

 [@CESSDA_Data](https://twitter.com/CESSDA_Data)



Mission of CESSDA

- ◇ Provide a distributed and sustainable **research infrastructure**
 - ◇ **enabling** the research community to conduct high-quality research in the social sciences,
 - ◇ contributing to the production of effective solutions to the **major challenges facing society today.**
- ◇ Facilitate **teaching and learning** in the social sciences.

Tools & services



Agenda

Introduction (Lisa Hirsch)

- ◊ What are Data Management Plans?
- ◊ Why are Data Management Plans so important?

How to write Data Management Plans (Anne Sofie Fink, Hannah Mihai)

- ◊ 2 examples

Q & A

What are DMPs?

- ◇ DMPS...
 - ◇ Are **living** documents
 - ◇ **Structure** and **systematize** your work with research data
 - ◇ View research data along the **research data life cycle**
 - ◇ Make data **FAIRer**

Why DMPs?

- ◇ Data Management Plans and **good Research Data Management** go hand in hand
- ◇ **Individual** benefits
- ◇ Requirements by **funders**
- ◇ **Plan** beforehand rather than during the research process to...
 - ◇ Allow for easy project management
 - ◇ Clarify the budget needed
 - ◇ Show accountability

How to write a DMP

- ◊ DMPs answer all the questions around the **provenance, storage, curation** and **access** to research data **during** and **after** your research project.



[CESSDA DMEG](#)

[List of DMP questions](#)

How to write DMPs?

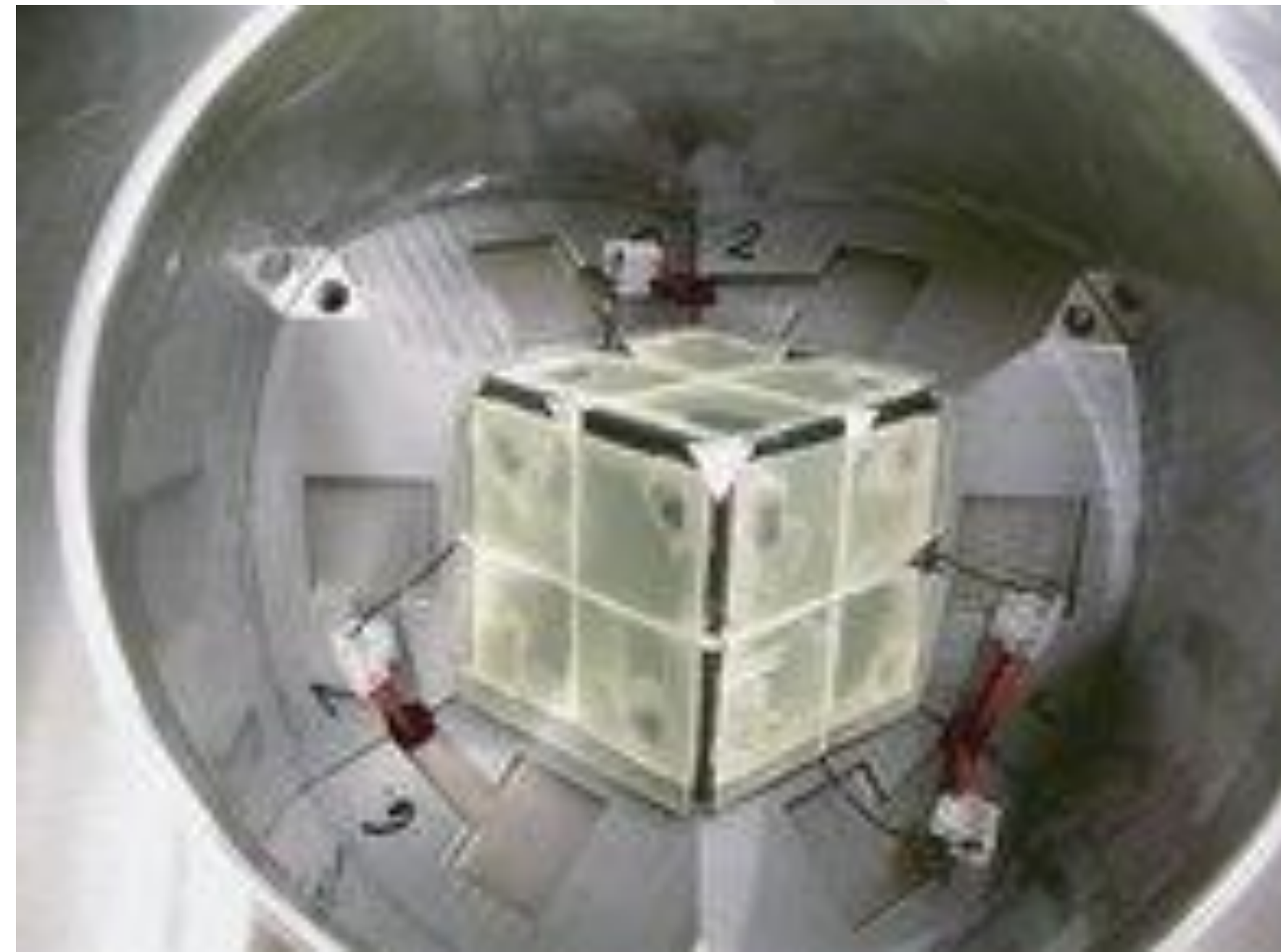
Anne Sofie Fink, Head of Data Management at DeIC
Hannah Mihai, Data Management Consultant at DeIC



Data Management is important...



<https://www.frm2.tum.de/en/frm2/news-single-view-en/article/the-new-multi-anvil-press-goes-into-operation/>



<https://www.gfz-potsdam.de/en/section/chemistry-and-physics-of-earth-materials/infrastructure/multi-anvil-press>

Data Management is important...

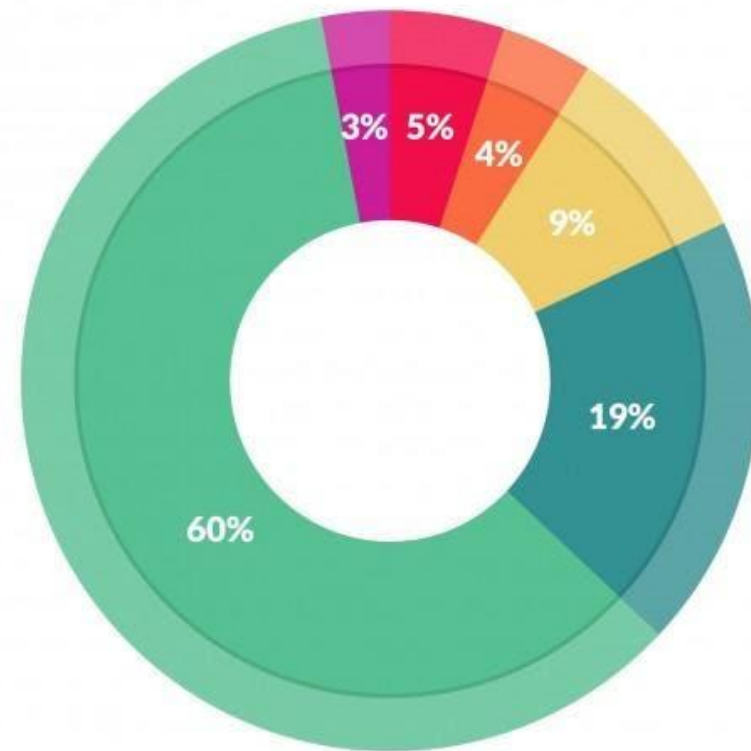
Table 3. Experimental Product Compositions for Melting Experiments at 1.5-2.3 GPa

Expt. Ph ^a	No.	SiO ₂	TiO ₂	Al ₂ O ₃	Cr ₂ O ₃	FeO	MgO	MnO	CaO	K ₂ O	Na ₂ O	P ₂ O ₅	Total	
L69	gl	12	45.4(2)	0.48(2)	16.3(1)	0.03(3)	8.95(14)	14.3(2)	0.06(3)	11.0(2)	0.08(2)	2.16(14)	0.11(4)	98.9
	ol	5	40.1(2)	0.01(1)	0.21(13)	0.01(1)	9.95(10)	48.2(4)	0.03(1)	0.30(5)	- ^b	-	-	98.8
	opx	26	54.0(5)	0.06(2)	8.26(74)	0.06(2)	6.13(7)	30.5(4)	0.04(2)	2.18(10)	-	0.15(14)	-	101.4
	cpx	8	52.0(4)	0.14(4)	9.39(5)	0.09(1)	4.89(16)	20.8(3)	0.06(2)	13.3(6)	-	0.69(5)	-	101.4
	gar	12	43.1(2)	0.12(4)	23.8(7)	0.14(3)	6.39(53)	21.7(3)	0.07(2)	5.29(16)	-	-	-	100.6

Run no.	Ph	SiO ₂	TiO ₂	Al ₂ O ₃	Cr ₂ O ₃	FeO	MnO	MgO	CaO	Na ₂ O	K ₂ O	Total
M5-103		51.64	0.09	7.17	0.60	4.97	0.10	24.57	10.13	0.71	0.02	
M5-40		48.53	0.52	12.37	0.12	9.02	0.20	16.64	10.89	1.65	0.06	
M7-16		43.58	0.75	13.73	0.07	14.51	0.30	12.52	13.77	0.75	0.03	

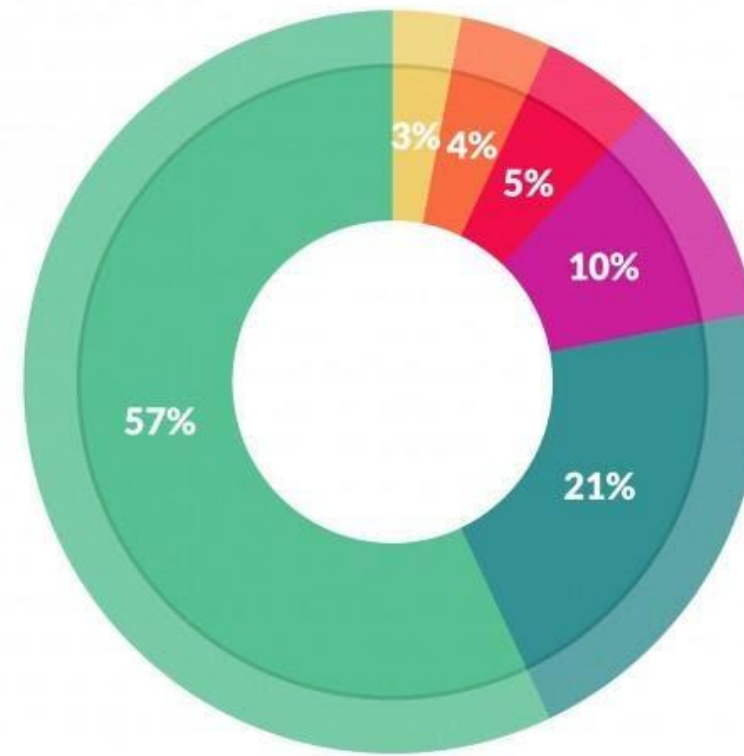
Run#	24 (20)	SiO ₂	49.53
SiO ₂	49.8 (4)	TiO ₂	0.51
TiO ₂	0.46 (8)	Al ₂ O ₃	16.28
Al ₂ O ₃	14.6 (2)	FeO*	7.54
Cr ₂ O ₃	0.28 (11)	MnO	0.15
FeO*	6.8 (2)	MgO	12.45
MnO	0.15 (8)	CaO	11.67
MgO	13.3 (2)	Na ₂ O	1.67
CaO	12.6 (2)	K ₂ O	0.07
Na ₂ O	1.58 (11)	Cr ₂ O ₃	0.13
K ₂ O	0.02 (2)	Total	100.00
Cl	<0.04		

Data Management is important...



What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets; 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%



What's the least enjoyable part of data science?

- Building training sets: 10%
- Cleaning and organizing data: 57%
- Collecting data sets: 21%
- Mining data for patterns: 3%
- Refining algorithms: 4%
- Other: 5%

Data Management is important...

- Because your research gets better!
- This means for yourself:
 - Increased chances of you understanding your own data after some weeks/months/years.
 - More data citations.
- This means for others:
 - Easier reuse of your data.
 - No need (or at least reduced need) to contact the author for clarification.
- Funder requirements
 - HE requires 2 DMPs in the course of the project
 - Webinar on HE requirements on DMPs: <https://deic.dk/da/event/webinar-14062023>
- [Data Sharing and Management Snafu in 3 Short Acts](#)
 - Can I have a copy of your data?
 - I have saved it on a USB drive and don't know where it is... And the program you need to open the file can't be downloaded anymore. What the column names mean? I don't know, I have created this data years ago, I can't remember...

The Horizon Europe DMP

- <https://enspire.science/wp-content/uploads/2021/09/Horizon-Europe-Data-Management-Plan-Template.pdf>
 - Data Summary
 - FAIR data
 - Making data findable, including provisions for metadata
 - Making data accessible
 - Making data interoperable
 - Increase data reuse
 - Other research outputs
 - Allocation of resources
 - Data Security
 - Ethics
 - Other issues



2 examples of DMPs

- SSHOC DMP by Anne Sofie
- Submerse DMP by Hannah



SSHOC DMP the OG and the follow up



- The [SSHOC Data Management Plan](#) (project starting)
 - purpose of the data collection in relation to SSHOC project objectives and activities
 - types and formats of SSHOC project data
 - reuse of existing data
 - origin of data and data usefulness
 - alignment with FAIR principles
 - resources needed and responsibilities within the project
 - data security
 - ethical and intellectual property aspects related to data.

SSHOC data handled throughout the project duration:

- **Survey data**
- **Case studies / pilots data**
 - **Tools and Service data**
 - **SSHOC Marketplace data**
- **SSHOC user communities data**
 - **Other data**

The update of the SSHOC DMP:

Overall Strategy for SSHOC Data Management



Ambitions for the DMP update

- Data management guidelines
- Best practices and ambitions for data management across the project for the last stages of the SSHOC project and beyond
- Documentation of argument for the continuous updates of the SSHOC DMP aka the SSHOC DMP as a living doc

Building blocks for the SSHOC DMP update

- List of all data sets and data usage
- FAIR Overview for all data sets
- Solution for sustainability and long-term preservation for all SSHOC metadata and data
- Keep the SSHOC DMP alive

- How data management planning has supported work in the SSHOC project
- Offers transparency to the SSHOC project's compliance with the FAIR principles
- How the project supports preservation and sustainability of data sets and data usages during and after the SSHOC project



Steps to improving the value of the SSHOC DMP



From DMP with data specificity to overview of SSHOC data

Action point 1: List of all datasets in one scheme



From specific descriptions for each data set to overall SSHOC data's FAIR accommodation

Action point 2: FAIR Overview for all data sets



Focus on sustainability and long-term preservation of all SSHOC metadata and data

Action point 3: Provide solution for sustainability and preservation more explicitly



Keeping the SSHOC DMP alive

Action point 4: Make a plan to update the SSHOC DMP



Ethics and GDPR – from data set unique to shared approach

Action point 5: Shared approach towards ethics and GDPR in SSHOC



Only data and metadata are covered by the DMP

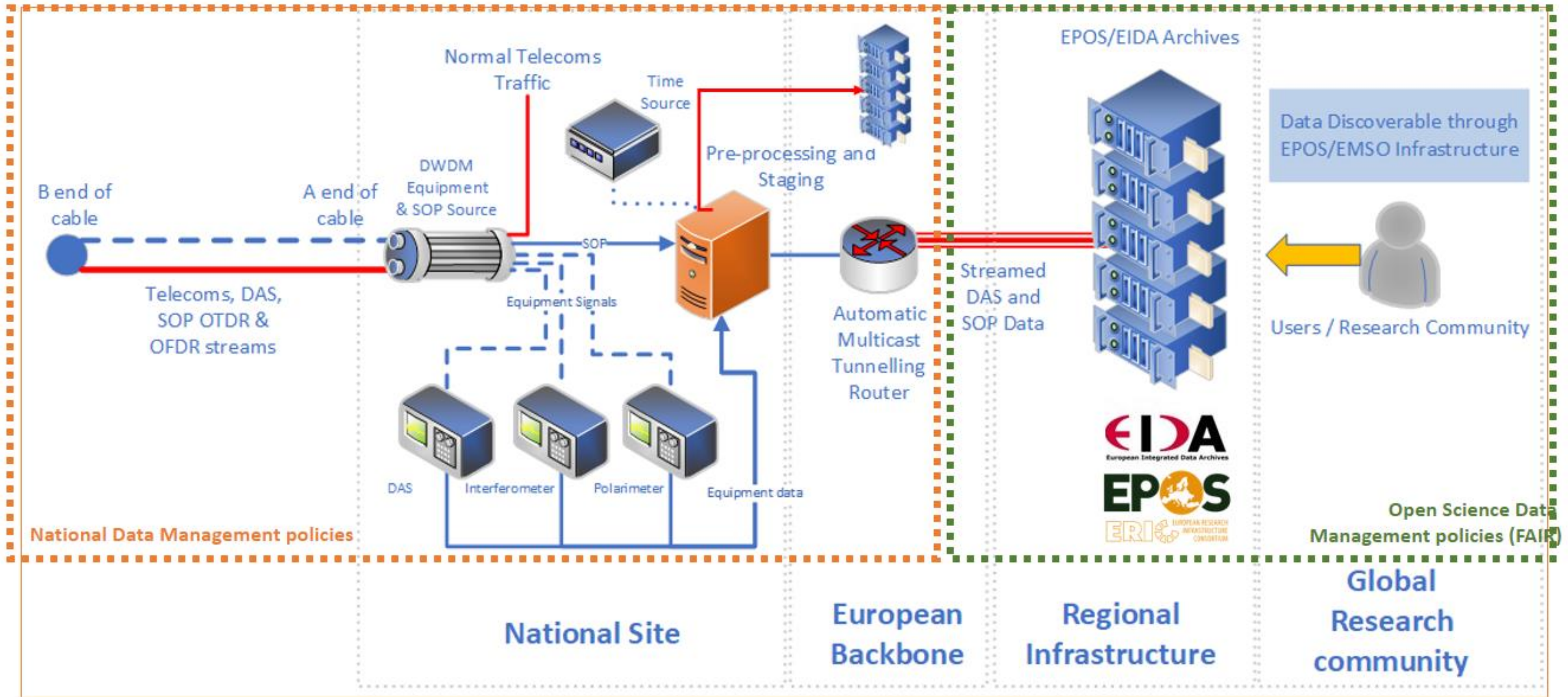
Action point 6: SSHOC DMP scope is data and metadata (as-is)

SSHOC DMP

- The FAIR Principles put to the front of the SSHOC DMP Follow-up
- Scoping the research data objects to plan for
- Start with dataset specificity and move to overview of data and metadata beyond the project
- Ethic & GDPR within/without
- DMP Sustainability and preservation
 - Re-use of data/digital research objects
 - Publications
 - Reproducibility of data
 - The overview of data incl. repository choice;)



Submerge DMP



Submerge DMP



Challenges generally about Data Management

The community needs to agree on

- where the data is published
- what standards to use for
 - Data formats
 - Metadata templates
 - Vocabularies

Challenges specifically about data collection on fiber-optical cables

- New data scrubbing workflows need to be created.
- Policies for data storage need to be created.
- Involvement of National Security Agencies.
- Satisfying a number of different research communities.
- Ensuring fast Open Data.
- What data should be stored / what should be deleted?
 - Someone's trash is someone else's data...

Q & A

Thank you for attending!