

# Further Professionalising Data Stewardship: engaging with and building on experience, the work of Research Data Alliance PDS Interest Group.

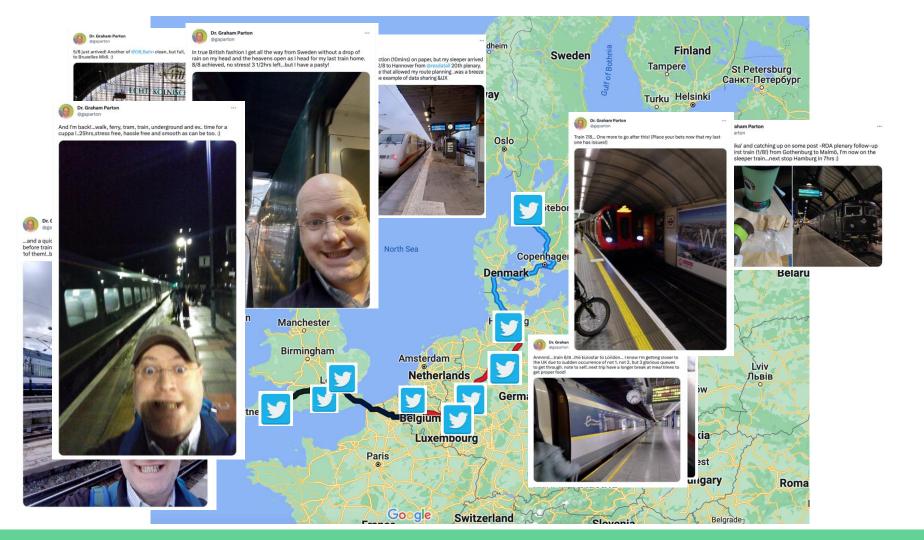
Graham Parton, Centre for Environmental Data Analysis | 0000-0003-4157-0352 Yan Wang, Delft University of Technology | 0000-0002-6317-7546 Jaana Pinnick, British Geological Survey | 0000-0002-7718-5133 Romain David, European Research Infrastructure on Highly Pathogenic Agents | 0000-0003-4073-7456

PV2023, May 2-4, 2023



# Introduction





But... more about that journey later...

...and a moment of realisation it sparked..

### The planning...

### ...relied on professional services that worked!

### February 2023



# RESEARCH DATA ALLIANCE in a nutshell

### Vision

### **Researchers and innovators**

openly share and **re-use** data across technologies, disciplines, and countries to address the grand challenges of society.

### **Mission**

RDA builds the social and technical bridges that enable open sharing and **re-use** of data.



www.rd-alliance.org

building the social and technical bridges that enable open sharing and re-use of data

64 FLAGSHIP	200+ ADOPTION
OUTPUTS	CASES
including 8 ICT	across multiple
Technical Specifications	disciplines,

#### 91 GROUPS WORKING ON GLOBAL DATA INTEROPERABILITY CHALLENGES

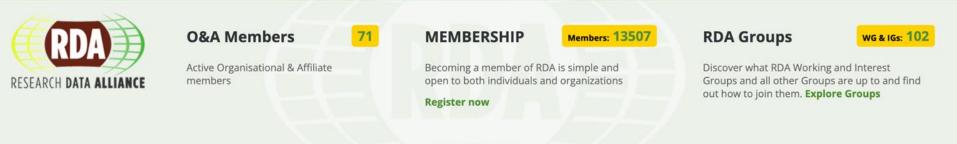
34 Working Groups 56 Interest Groups 1 Community of Practice

#### 13,338 INDIVIDUAL MEMBERS FROM 150 COUNTRIES

69% Academia & Research 14% Public Administration 11% Enterprise & Industry

#### 61 ORGANISATIONAL MEMBERS 12 AFFILIATE MEMBERS





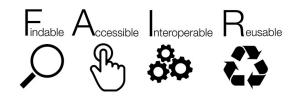
- over 35 regional networks covering 148 countries
- builds the social and technical infrastructure to enable open sharing and reuse of data.
- has a grass-roots, inclusive approach covering all data lifecycle stages, engaging data producers, users and stewards, addressing data exchange, processing, and storage.





### (Research) Data Stewardship Definition:

*"Responsible planning and executing of all actions on (digital) data before, during and after a research project, with the aim of optimizing the usability, reusability and reproducibility of the resulting data"* (source: <u>DTL</u>)





## Data stewardship and data management skills are essential in the research process.....BUT...

"[a] lack of consensus on the responsibilities, knowledge, skills, and value of data stewards ... leads to confusion about the role and hampers building adequate data steward capacity... [C]ombined with scarce resources for data stewardship [...this] limits vital developments such as defining recruitment procedures and career paths for current and future data stewards."

**RDA PDS-IG charter** 

## And that required capacity is huge!



Mons estimated at least 10 million serious data producers among the EU's 70 million science and technology professionals and 1.7 million researchers, arguing a need to:

"...educate about **500,000 data stewards** of various kinds to support researchers through experimental design and data capture, curation, storage, analytics, publication and reuse." ...

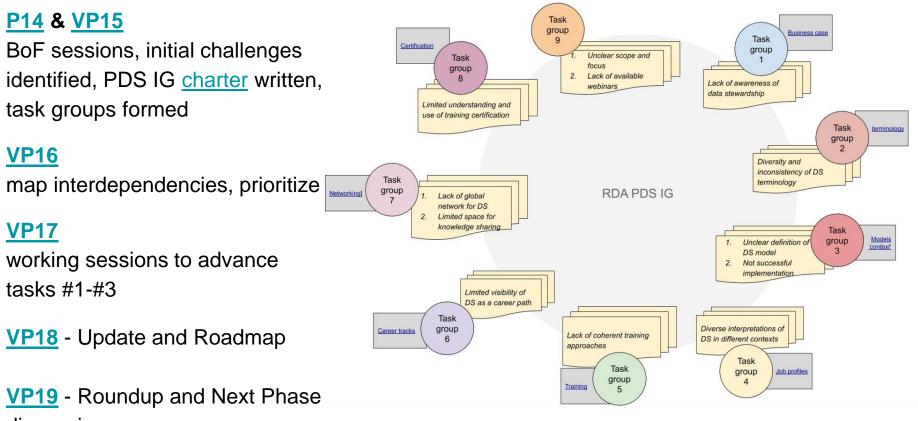
...and for the European Union *"with* €300 *billion (US*\$325 *billion) of public money spent on research* ... should expect to spend €15 billion on data stewardship."

Barend Mons (2020)

### 2020 : RDA Professionalising Data Stewardship IG constituted

Building the social and t	technical bridge	s to enable open sha	aring and re-use of da	ata	RDA EU	RDA US CONTACT	US LOGIN REGIS	TRATION	ツ 🕞 🗄 🛄 🗎
CREATE ALLIANCE O&A Members 71 Active Organisational & Affiliate members			Bec	MEMBERSHIP Becoming a member of RDA is simple and open to both individuals and organizations Register now			RDA Groups we & IGs: 102 Discover what RDA Working and Interest Groups and all other Groups are up to and find out how to join them. Explore Groups		
ABOUT RDA * GET INVOLVED * GROUPS * RECOMMENDATIONS RDA FOR DISCIPLINES * PLENARIES & EVENTS * NEWS & MEDIA * & OUTPUTS * Home » Working and Interest Groups » Interest Group » Professionalising Data Stewardship IG Professionalising Data Stewardship IG Taxonomy:									
Posts C	(reate Wiki index	Events	Repository	Outputs	Charter	Plenaries	Members	create new	r content ▼∽
Group Status: 📀	IG Establish	ned						8+	Join Group
Please make sure t questions.	the group fo	llows the new R	DA Groups Polic	<b>y</b> , which came ir	nto effect on 1 Ap	ril 2021. Please e	contact enquirie	es[at]rd-alliance.org i	f you have any

## RDA PDS IG History - From Helsinki in 2019 to PV2023



discussions

## Work to date









Romain... I think I'm honing in on a concept for the PV paper....

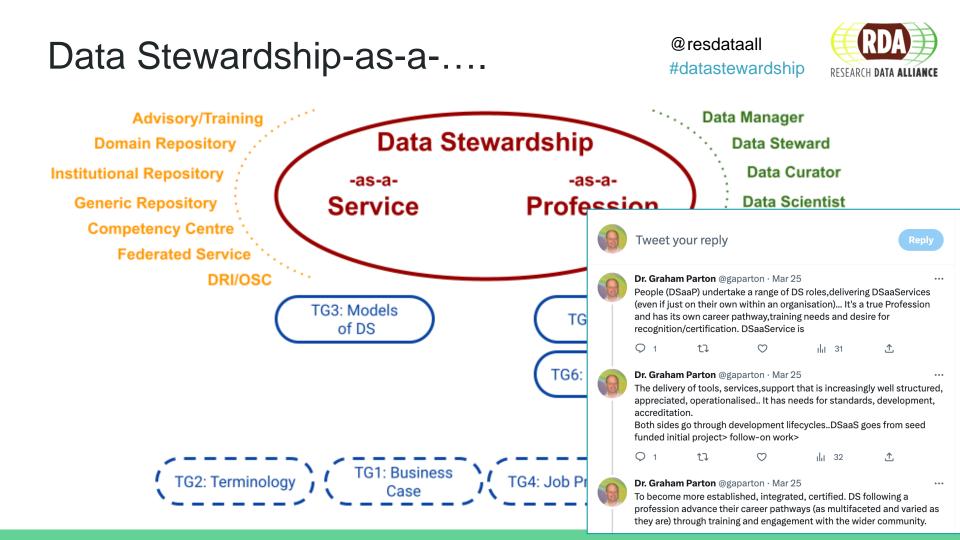
DSaaS and DSaaP .... Which is ..DataStewardshipas-a-Servce. & DataStewardship-as-a-Profession

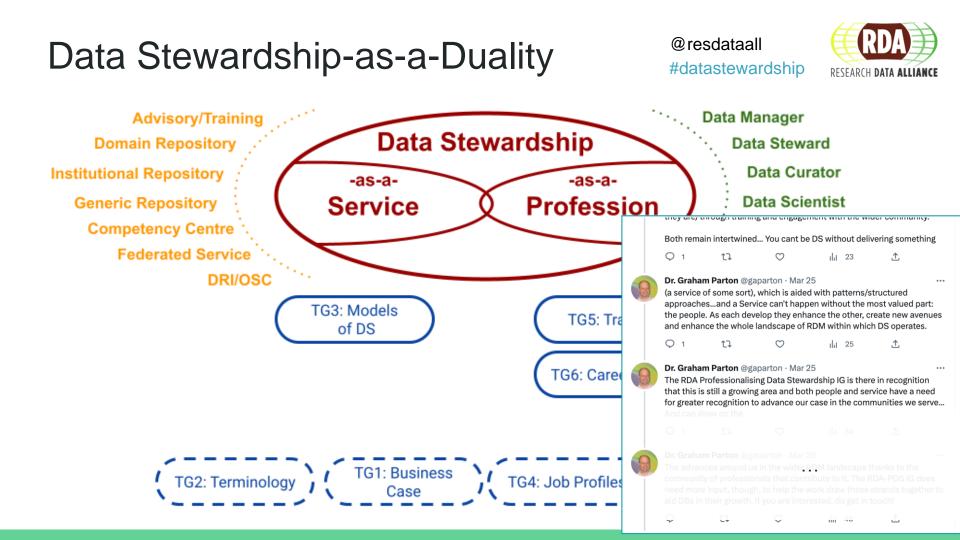
One is service orientated, the other people orientated, but both go hand in hand. You cant deliver a service without poeple ...and the people deliver the service. ... It maps well to our IG structure and progress-pathways on both sides ... Services need business cases, evolve and we can link to guidence elsewhere in RDA on what to consider to evolve this way... Whilst people need training, career pathways with defined, understood roles

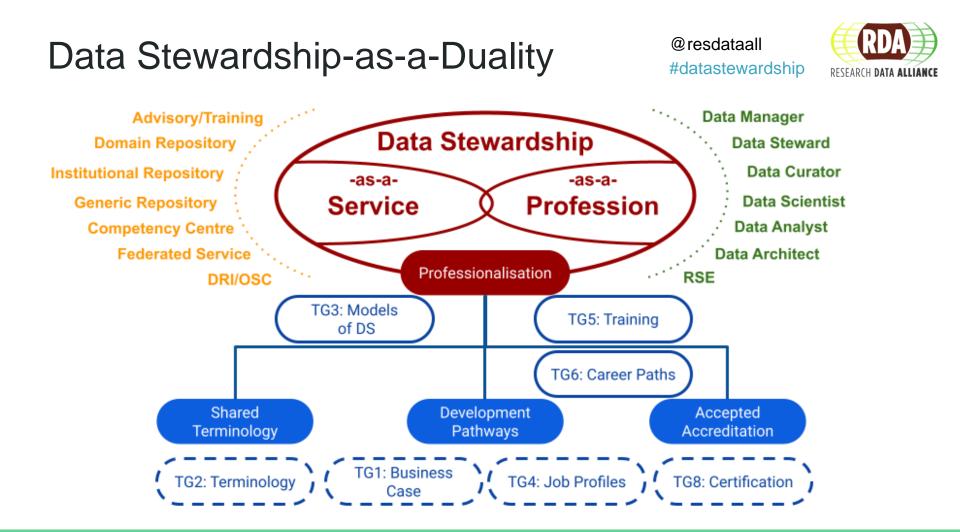
Mar 24, 2023, 7:59 PM

It plays on the SAAS ,PasS, CaaS paradigms that will be familiar with others too and hopefully draw them in.

Aunich Austria Budapest Hungary Roma









# The DSaaS focus: Models of Data Stewardship



# Group 3: Data Stewardship Models

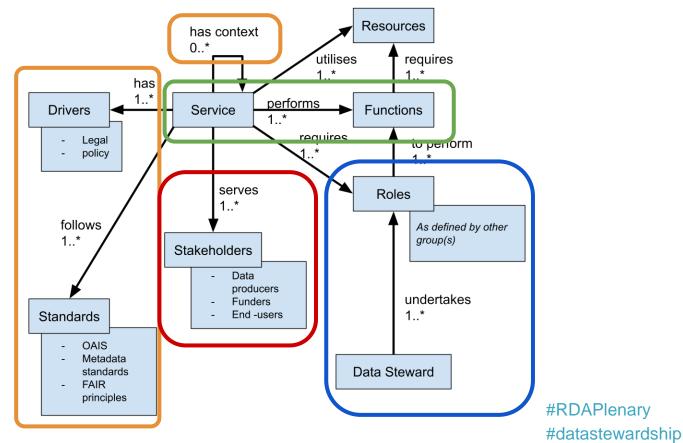
Graham Parton (co-chair) Bill Ayres (co-task group lead) Liise Lehtsalu (co-task group lead) Niklas Zimmer, Angus Whyte, Iryna Kuchma, Adam Száldobágyi, Eleanor Warren

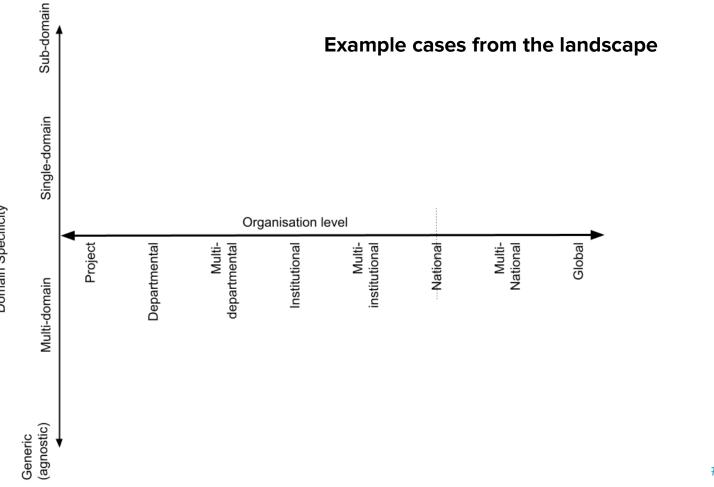
## Conceptual diagram and TG aims



We aim to model the Data Stewardship Roles within particular service provisions of Data Stewardship, the larger Organisational Contexts within which such provisions exist, and how those roles

such provisions exist, and how those roles and services interact with their **Target Communities**.



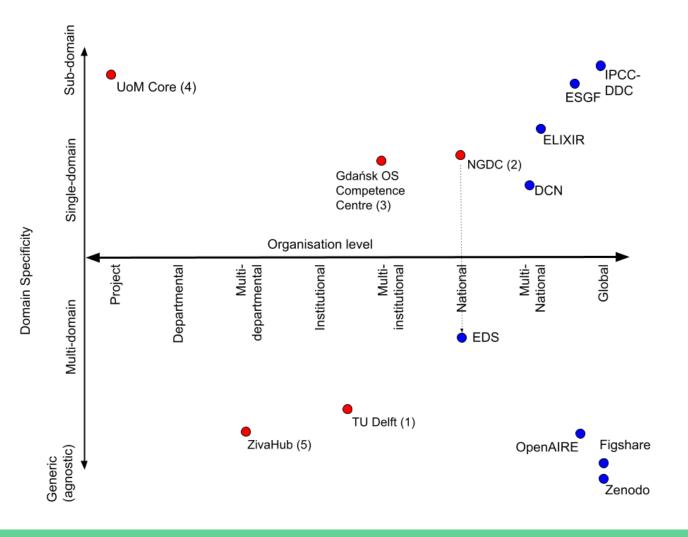


Domain Specificity

#RDAPlenary #datastewardship

RESEARCH DATA ALLIANCE

GP

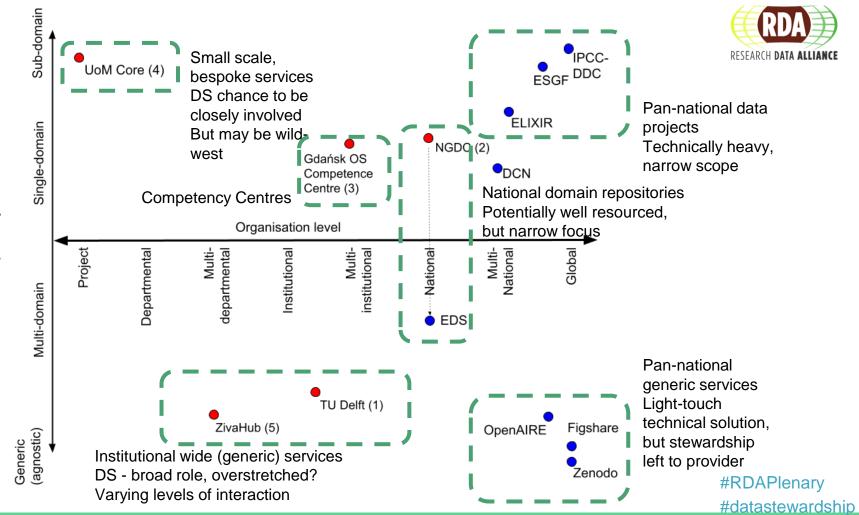


**#RDAPlenary** 

#datastewardship

RESEARCH DATA ALLIANCE

GP



Domain Specificity

GP

## Survey report and dataset available in Zenodo



#### DOI: 10.15497/RDA00075

December 14, 2022

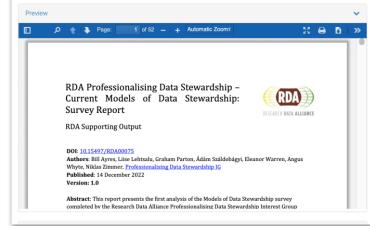
Other Open Access

#### RDA Professionalising Data Stewardship -Current Models of Data Stewardship: Survey Report

🔞 Ayres, Bill; 💿 Lehtsalu, Liise; 🙆 Parton, Graham; 💿 Száldobágyi, Ádám; 🕲 Warren, Eleanor, 🔞 Whyte, Angus; 💿 Zimmer, Niklas

This report presents the first analysis of the Models of Data Stewardship survey completed by the Research Data Alliance Professionalising Data Stewardship Interest Group (RDA PDS-IG) in October-November 2021. The purpose of the survey, based around a conceptual model of data stewardship services, was to engage with the global research data management (RDM) community to capture information about data stewardship approaches taken by organisations and the wider landscape of data stewardship and services. In particular, this initial report provides insights on a) roles involved in providing data stewardship services, the functions these provide, to which customer groups, and at which stages of the data lifecycle; b) stakeholder involved in the governance of data stewardship; c) indicative ranges of service capacity of data stewardship services. It data stewardship promoted by services; and e) main lessons learned by responding data stewardship services.

The dataset collected in the survey can be viewed at doi.org/10.5281/zenodo.6665306



June 19. 2022

#### DOI: 10.5281/zenodo.6665306



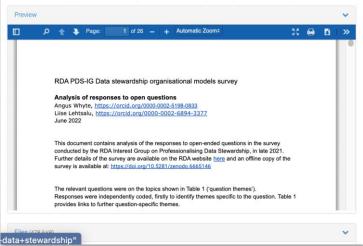
#### RDA Data Stewardship Organisational Models Survey 2021 Output Dataset

🔞 Ayres, Bill; 🕲 Lehtsalu, Liise; 🚱 Kuchma, Iryna; 🕲 Parton, Graham; 🕲 Száldobágyi, Ádám; 🕲 Warren, Eleanor; 🙆 Whyte, Angus; 🔞 Zimmer, Niklas

Data Stewardship comes in many forms and contexts, with the common goal of supporting data management. Yet that diversity can make it hard for the RDM community to further professionalise our work and services.

The RDA Professionalising Data Stewardship Interest Group (PDS-IG) Models Task Group sought input from the research data community to help model different approaches to research data stewardship through an online survey in October to November 2021. An offline copy of the survey is available as a linked resource from this record.

This dataset consists of a cleaned, anonymised responses from 136 respondents, though not all questions were answered by all respondents. The output are available in a CSV formatted file containing the full response set along with output from a thematic textual analysis undertaken on responses to a number of open-text questions in the survey.





# The DSaaP focus: Career Pathways and Training Task Groups



# Group 5: Research landscape of data stewardship training

Authors: Jukka Rantasaari 0000-0001-5927-3781 Lorna Wildgaard 0000-0002-3900-5058 Editors: Claudia Engelhardt 0000-0002-3391-7638 Jaana Pinnick 0000-0002-7718-5133



Gaps in data stewardship: What kind of needs for training do data stewards face in supporting research?

### Take home messages:

- 1. Collaborate internationally on certified training
- 2. Enable researchers to share their data from the planning stage of their project
- 3. Address the increased need for technical skills in data stewardship

RESEARCH DATA ALLIANCE

Gaps in data stewardship: What kind of needs for training do data stewards face in supporting research?



Picture credit: Patrick Hochstenbach in https://zenodo.org/record/1212496

Authors Jukka Rantasaari,0000-0001-5927-3781 Lorna Wildgaard, 0000-0002-3900-5058

Editors Claudia Engelhardt, 0000-0002-3391-7638 Jaana Pinnick, 0000-0002-7718-5133

## Landscape report and Matrix available in Zenodo

Report Open Access

22 🖨 🖬 |

**>>** 



#### October 4, 2022

DOI: 10 15497/RDA00076

RDA Professionalising Data Stewardship - Data Stewardship Landscape Initial Report

Wildgaard, Lorna; Rantasaari, Jukka

#### Editor(s)

Preview

Engelhardt, Claudia; Pinnick, Jaana

This report is based on a literature review published by the Research Data Alliance Professionalising Data Stewardship Interest Group (RDA PDS-IG) Training sub-group in October 2022, A framework for the literature collection was devised before the literature search began, based on common concepts of the division of the research disciplines. The literature search was conducted between March 2021 and July 2021. The selection was verified in a cross-walk exercise, where common terminologies were tracked across different ontologies to learn more about concepts and disciplinary differences and commonalities in research data management (RDM). In the reference management software Zotero, folders were created to structure the literature into the following categories: data stewardship; education and training of data stewards; RDM education; training materials; data management plan templates; documents about the skills needed to support EOSC and FAIR principles; data management in the humanities, industry, life sciences, natural sciences, public institutions and the social sciences. Further, the aim was to identify papers concerning the RDM needs of researchers and data stewards and their role in the RDM life cycle. The screening resulted in 125 papers that are used as the knowledge base for this report.



DESCADOU DATA ALLIANO

Gaps in data stewardship: What kind of needs for training do data stewards face in supporting research?

3 of 79 - + Automatic Zoom=





#### RDA Professionalising Data Stewardship - Data Stewardship Landscape Report Resource Matrix

#### Wildgaard, Lorna; Rantasaari, Jukka

This dataset includes the resources which were included in the Research Data Alliance Professionalising Data Stewardship Interest Group (RDA PDS-IG) Training Landscape sub-group report published in October 2022. The literature was structured into the following categories; data stewardship; education and training of data stewards; RDM education; training materials; data management plan templates; documents about the skills needed to support EOSC and FAIR principles; data management in the humanities, industry, life sciences, natural sciences, public institutions and the social sciences. The aim was to identify papers concerning the RDM needs of researchers and data stewards and their role in the RDM life cycle The screening resulted in 125 papers that are used as the knowledge base for the output report

#### Preview

October 4, 2022

#### 1 of 10 - + Automatic Zoom: 🐌 Page:

RDA Professionalising Data Stewardship -Data Stewardship Landscape Report **Resource Matrix** 



**RDA Supporting Output** 

DOI: 10.15497/RDA00077

Authors: Lorna Wildgaard, Jukka Rantasaari

Published: 4 October 2022

Version: 1.0

Name

Shannan M. M. L. Jakanak India da aka mananana sakish sama india da dia da M. Mananah Mar

230 1 kR

#### Files (230.1 kB) Size

andecane Report Matrix Final Zenodo pd

Proview
 Downloo

~

X 🖨 🖪



# Group 6: Data Steward Career Tracks

Biru Zhou 0000-0001-6914-1432 Liise Lehtsalu 0000-0002-6894-3377 Valerie Mccutcheon 0000-0002-0705-4832 Elizabeth Newbold 0000-0002-8255-9013 Yan Wang 0000-0002-6317-7546

## Data Stewardship as a profession in academia

### Challenges

- Research Data Management professional like Data Steward has not been recognized as a profession in academia
- Uncertain career paths, unsustainable pool of expertise, etc.

### **Group plan**

• Community survey to understand current and future career perspectives

### **Expected outcomes**

- Reach out to people who is/was working as Data Steward or similar RDM support roles across different regions and organizations
- Initial overview on the career perspective from the community
- Seek for volunteers who would like to provide more insightful inputs as case studies



## **Data Steward Career Tracks Survey**

- Ran in October 2022
- Report due soon!

## Next steps – webinar series with EOSC TF

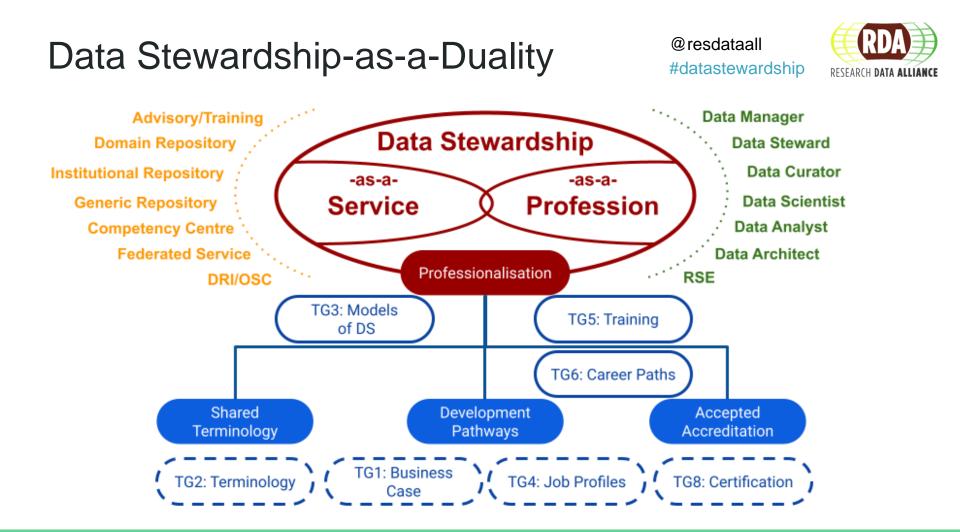
Define a range of personas

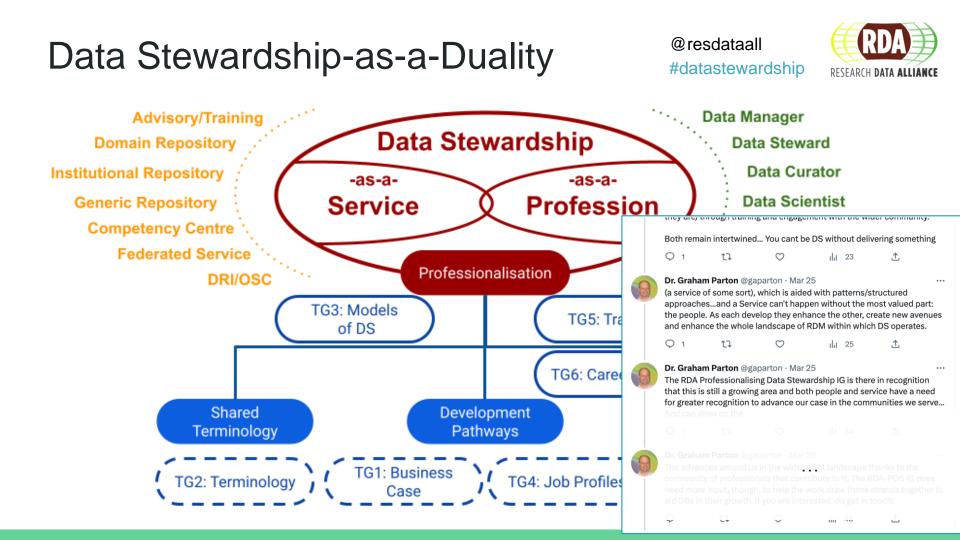


### EUROPEAN OPEN SCIENCE CLOUD



# Other aspects and Future Work: Furthering the Professionalisation







# Connecting with, and contributing to, the work of the PDS-IG

# Next steps - Keeping in touch

- Up to the next plenary (P21, IDW 2023, 23-26 October 2023, Salzburg, Austria)
  - Participate / continue working in the task groups
  - Help formulate and participate in the Working Group proposals
  - Join our monthly meetings

We need you to make professionalising data stewardship a real community effort, by and for the community!

- So, you are very welcome to join us
- Contact your group's contact person to become actively involved
- Make sure to check into our <u>slack community</u> for regular communication
- Some TGs have regular Zoom meetings contact them to get link
- Questions? Ideas? Want to collaborate? Make sure to contact us





# https://www.rd-alliance.org/groups/professionalising-data-stewardship-ig

Building the social and technical b	lata	RDA EU RDA US CONTACT US LOGIN			RATION	y 🕞 🗄 🛄 🗎	
RESEARCH DATA ALLIANCE O&A Members 71 Active Organisational & Affiliate members			MEMBERSHIP Becoming a member of RDA is simple and open to both individuals and organizations Register now			RDA Groups wG & IGS: 102 Discover what RDA Working and Interest Groups and all other Groups are up to and find out how to join them. Explore Groups	
ABOUT RDA 🎽 GET INVOL	VED * GROUPS * RECOMMEND & OUTPUTS *		R DISCIPLINES 👻	PLENARIES & EV	/ENTS * NEW	∕S & MEDIA Ÿ	٩
Home » Working and Interest Groups » Interest Group » Professionalising Data Stewardship IG Professionalising Data Stewardship IG Taxonomy:							
Posts Create Wiki	Events Repository	Outputs	Charter	Plenaries	Members	create new	content V
Group Status: 🐼 IG Estal	blished					8.	Join Group
Please make sure the grou questions.	o follows the new RDA Groups Poli	<mark>cy</mark> , which came in	to effect on 1 Ap	il 2021. Please co	ontact enquirie	s[at]rd-alliance.org if	you have any



# Thanks and Acknowledgements

Past co-chairs, members of the various Task Groups and the Interest Group community of over 270 members!

# Past Co-Chairs of the PDS IG





Mijke Jetten, Dutch Techcentre for the Life Sciences (DTL), Data Stewardship Community Manager, mijke.jetten@dtls.nl



Debora Drucker, Brazilian Agricultural Research Corporation (Embrapa), debora.drucker@embrapa.br



Peter Neish, Melbourne University, Program Manager Stewardship & Open Research, peter.neish@unimelb.edu.au



Varsha Khodiyar, Springer Nature Data Curation Manager, varsha.khodiyar@nature.com



<u>Christina Drummond</u>, Educopia Institute Data Trust Program Officer, christina@educopia.org

# Present Co-Chairs of the PDS IG





Jaana Pinnick, British Geological Survey/ National Geoscience Data Centre, Grants Manager, jpak@bgs.ac.uk



<u>Graham Parton</u>, Centre for Environmental Data Analysis (CEDA), Senior Environmental Data Scientist, graham.parton@stfc.ac.uk



Romain David, European Research Infrastructure on Highly Pathogenic Agents, Research fellow, data manager, romain.david@erinha.eu



Niklas Zimmer, University of Cape Town, Manager Digital Library Services, niklas.zimmer@uct.ac.za



Yan Wang, Delft University of Technology, Data Stewardship Coordinator, y.wang-16@tudelft.nl



# Appendix 1 Task group challenges, steps, expected outcomes & RDA input requests

# Group 1: Business case



# Challenges

• Organisations rely on using data for research and analysis, but are often not aware that they need somebody with data stewardship expertise

# Steps

- Case studies to highlight the benefits of data stewardship to stakeholders
- Risk assessment: what might go wrong if data stewardship expertise is lacking

## **Expected outcomes**

• Short report with case studies and stakeholder mapping exercise, to be used as a business case for data stewardship

# Contact person: <u>debora.drucker@embrapa.br</u>

# Task group 2: Terminology



#### Challenges

- Combining terminologies in data stewardship to bring different organisations together
- Highlighting polysemic issues and various uses of terms across domains

### Steps

- Re-using existing work from RDA and other fora, and enrich existing tools
- Interactions with other TG to find and propose the definitions they need
- Formalising a community approval process to choose the better definitions and vocabularies

#### **Expected outcomes**

- A data stewardship terminology tool to share and validate terms
- A first version of a minimal set of common terms definitions

### Input from the RDA community at the upcoming plenary session

- Feedback and suggestions: are we taking the correct approach?
- Is there any work or vocabularies that we should also consider?

### Contact person: romain.david@erinha.eu

# Task group 3: Models



#### Challenges

• Ill-defined models for data stewardship service provision, which may lead to duplication of efforts and the implementation of inappropriate strategies

#### Steps

• An environmental scan of data stewardship and governance staffing models, using the open business model canvas methodology, to develop a representative collection of real-world, reusable models

#### **Expected outcomes**

• Open available collection of models, to serve in establishing and sustaining data stewardship services at different organisations

#### Input from the RDA community at the upcoming plenary session

• Advice on existing materials and definitions created by other IGs to reduce duplication of efforts

Contact person: niklas.zimmer@uct.ac.za

# Task group 4: Job profiles



### Challenges

• Understanding the data steward role in different working environments and contexts

### Steps

- Collecting job profiles of data stewards and of other roles explicitly mentioning data stewardship in the role profile and responsibilities
- Reviewing existing and previous work on this topic

### **Expected outcomes**

- A library of example job profiles
- A matrix of commonalities and differences in job profiles

### Input from the RDA community at the upcoming plenary session

- Job descriptions of RDA members in data steward and data-related positions
- References to previous work or projects of relevance

Contact person: <a href="mailto:varsha.khodiyar@nature.com">varsha.khodiyar@nature.com</a>

# Task group 5: Training



#### Challenges

• Limited coherent approach, making the global demand for data stewards challenging

### Steps

- Contact and collaboration with initiatives having similar aims
- Survey existing literature, tools and lists using RDA terminology and models

### Expected outcomes

- Gap analysis to identify the gaps in data stewards' skills, researchers' skills, and in existing training materials
- By identifying gaps and relevant training materials we will advise the creation of training pathways (e.g. in <u>Competency Hub (ebi.ac.uk)</u>)

### Input from the RDA community at the upcoming plenary session

- Sources (academic publications) where we can find answers to the questions in the slide 18?
- Please come and join us

#### Contact person: Jaana Pinnick jpak@bgs.ac.uk

# Task group 6: Career tracks



#### Challenges

• Limited visibility of data stewardship as a career path: what do new recruits need on entering the profession and how might they progress in their roles?

#### Steps

- Interviewing data stewards about how they got into the profession and their career development
- Collating existing published profiles and articles about data stewards

#### **Expected outcomes**

 Case studies and composite personas of data stewards at different stages in their careers in different sectors

#### Input from the RDA community at the upcoming plenary session

- Feedback on whether this outcome will be useful
- Feedback on which sectors should be targeted (eg. academic, commercial, community)
- Volunteers to be interviewed or suggestions on how to recruit interviewees

#### Contact person: y.wang-16@tudelft.nl

# Task group 7: Networking



#### Challenges

• Explore the need for a global data stewards network, for sharing of knowledge and practices

#### Steps

- Collection of case studies and interviews with data stewards and related communities of practice
- Investigate models of community formation and practice exchange

#### **Expected outcomes**

- Synthesise information gathered through case studies and interviews
- As a first step, initiate quarterly informal online gatherings

### Input from the RDA community at the upcoming plenary session

- We need more task group members
- Watch for opportunities to contribute or take part in networking events

Contact person: <a href="mailto:peter.neish@unimelb.edu.au">peter.neish@unimelb.edu.au</a>

# Task group 8: Certification



#### Challenges

• Explore the use of and need for certification of training programmes and their data stewards

### Steps

- Collaborate with existing (industry/research) accreditation frameworks and badging systems
- Investigate lightweight certification, such as gamified badging
- Investigate how to recognise skills and expertise of data stewards

#### **Expected outcomes**

- Self assessment checklist for data stewards, including levels (beginner, intermediate, advanced)
- Recommendations for certification and badging systems for training and data stewards

### Input from the RDA community at the upcoming plenary session

• Invite people with certification expertise from research and industry to join the task group

Contact person: mijke.jetten@dtls.nl

# Task group 9: Webinars - on Hold since VP16



### Challenges

- Identify the audience, value, and need for webinars among research data stewards
- Assess if webinars are already provided elsewhere (e.g. via training and networking TG)

### Steps

Coordinating with other IG task groups and members

### Expected outcomes

• Recommendations for webinar series topics and presenting organisations/experts

## Input from the RDA community at the upcoming plenary session

- Ideation around where data stewards currently get webinar information and what's missing
- Challenge: lacking volunteer webinar leaders / webinar logistics and administration
- Question: is this task really a sub-task related to training and networking?
- Question: where can we have the most impact without duplication of effort?

Contact person: christina@educopia.org