



# Il ruolo e le competenze del “data steward”

Una nuova figura professionale a supporto dei dati FAIR per la ricerca

29 ottobre – GenOA Week 2021

Valentina Pasquale, PhD  
Istituto Italiano di Tecnologia



Quest'opera è distribuita con Licenza [Creative Commons Attribuzione 4.0 Internazionale](https://creativecommons.org/licenses/by/4.0/).

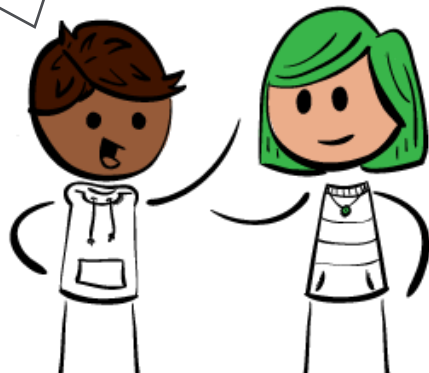
# Che cosa succede quando i dati non sono gestiti?

## I dati non sono più riutilizzabili!!!

### Mancanza di documentazione

Steve left the lab 2y ago. I would like to use his PhD data, but I can't find the HD where he saved them!

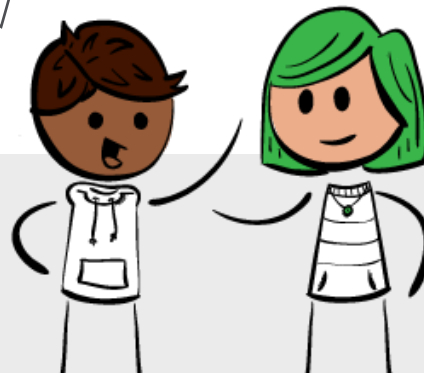
He could have used a repository to preserve data for the long-term!



Perdita dei dati

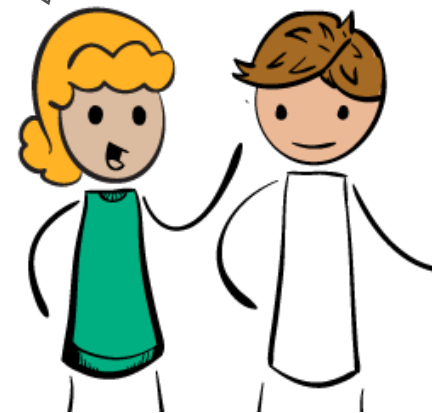
After a while....

I managed to find Steve's data and the boss asked me to analyze them, but I don't know what "var1" and "var2" mean!



I wish to reproduce the findings of that old paper by Prof. Rossi, so I've asked him the data. Unfortunately, the software needed to open it is no longer available!

He should have converted it to an open interoperable format!



Obsolescenza dei formati e degli strumenti software

# Senza i dati, la scienza è...

## Non riproducibile e poco trasparente

Senza la disponibilità dei dati a supporto delle pubblicazioni, gli studi non possono essere riprodotti e sottoposti a verifiche indipendenti (principio alla base del metodo scientifico...).

Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016). <https://doi.org/10.1038/533452a>

## Inefficiente

I dati della ricerca possono essere utilizzati per rispondere a più di un quesito scientifico: la loro condivisione e corretta gestione abilita il riutilizzo futuro, sia per gli autori sia per altri ricercatori, evitando duplicazioni.

U. von der Leyen's speech at #davos2020 (22nd January 2020) [https://www.youtube.com/watch?v=\\_A7Q514z\\_dw](https://www.youtube.com/watch?v=_A7Q514z_dw)

## Di minore impatto sociale ed economico


La scienza aperta moltiplica l'impatto della ricerca a tutti i livelli della società, risponde alla necessità di rendere conto ai cittadini dei finanziamenti pubblici alla ricerca, contribuisce al benessere collettivo e globale, ispirandosi ai valori di equità e inclusività.

[UNESCO Recommendation on Open Science](#)

# Dati FAIR e dati open

I principi FAIR nascono nel 2016 come linea guida per una corretta gestione dei dati della ricerca.

<https://www.go-fair.org/fair-principles>

**Box 2 | The FAIR Guiding Principles** 

**4 principi fondamentali**  
**15 principi accessori**

**To be Findable:**  
F1. (meta)data are assigned a globally unique and persistent identifier  
F2. data are described with rich metadata (defined by R1 below)  
F3. metadata clearly and explicitly include the identifier of the data it describes  
F4. (meta)data are registered or indexed in a searchable resource

**To be Accessible:**  
A1. (meta)data are retrievable by their identifier using a standardized communications protocol  
A1.1 the protocol is open, free, and universally implementable  
A1.2 the protocol allows for an authentication and authorization procedure, where necessary  
A2. metadata are accessible, even when the data are no longer available

**To be Interoperable:**  
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.  
I2. (meta)data use vocabularies that follow FAIR principles  
I3. (meta)data include qualified references to other (meta)data

**To be Reusable:**  
R1. meta(data) are richly described with a plurality of accurate and relevant attributes  
R1.1. (meta)data are released with a clear and accessible data usage license  
R1.2. (meta)data are associated with detailed provenance  
R1.3. (meta)data meet domain-relevant community standards

F indable

A ccessible

I nteroperable

R eusable

Per gli esseri  
umani e per le  
macchine

F ully

A (rtificial)

I (ntelligence)

R eady

La ricerca si basa sempre di più sul supporto computazionale delle macchine per trattare ed analizzare i dati, in risposta all'aumento del **volume**, della **complessità** e della **velocità di creazione** dei dati.

I **principi FAIR** enfatizzano non solo la riutilizzabilità dei dati da parte degli esseri umani, ma anche la capacità di **trovare, accedere, combinare** e **riutilizzare** i dati attraverso l'uso di strumenti computazionali (algoritmi) con minimo intervento umano.

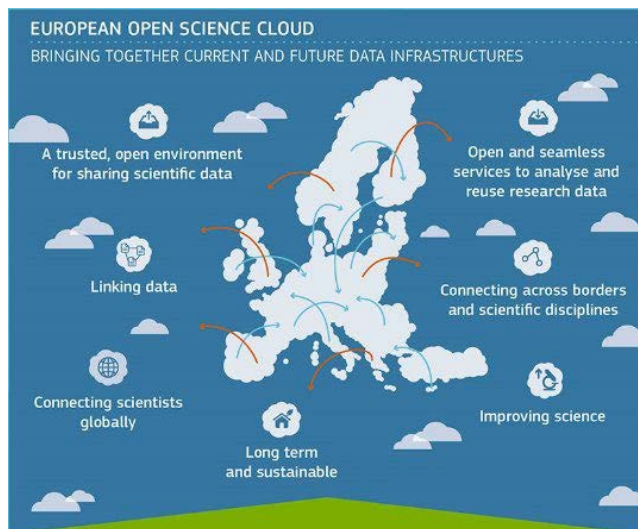
Sorgente: sito web GO FAIR.

## FAIR data ≠ open data

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

Valentina Pasquale – Istituto Italiano di Tecnologia

# EOSC: il web dei dati FAIR



U. von der Leyen's speech at #davos2020 (22<sup>nd</sup> January 2020)  
[https://www.youtube.com/watch?v=\\_A7Q514z\\_dw](https://www.youtube.com/watch?v=_A7Q514z_dw)

**We are creating a European Open Science Cloud now.**  
It is a trusted space for researchers to store their data and to access data from researchers from all other disciplines. [...]

Every researcher will be able to better use not only their own data, but also those of others. They will thus come to **new insights, new findings and new solutions.** [...]



La **European Open Science Cloud** contribuirà a realizzare l'Internet dei dati e dei servizi FAIR (in inglese, "**Internet of FAIR Data and Services**"), un ecosistema di servizi, norme, linee guida, e strumenti che abiliteranno la produzione, la condivisione e il riuso di "FAIR digital objects".

# Perché servono i « data steward»?

EOSC SRIA v1.0

Per poter produrre dati FAIR i ricercatori hanno bisogno di un supporto professionale adeguato, così come già avviene in altri contesti (TT, comunicazione, etc.)

Strategic Research and Innovation Agenda  
(SRIA)  
of the  
European Open Science Cloud (EOSC)

Version 1.0 15 February 2021

News Blog

## ← We need 500.000 respected data stewards to operate the European Open Science Cloud [Interview to Barend Mons, Chairman of the High Level Expert Group on EOSC, May 2016](#)

04/05/16 09:08

At the e-IRG workshop in Amsterdam, we had the opportunity to talk to Barend Mons who is chairing the High Level Expert Group on the European Open Science Cloud, an advisory group to the European Commission. To be successful, the European Science Cloud needs a lot of experts to operate it. Barend Mons told us. Data stewards that have a lot of knowledge about managing and maintaining data. Experts who are well respected with a solid career path. Barend Mons also discussed several other findings of the Expert group, whose report will be published very soon.



We are here at the e-IRG workshop in Amsterdam and we are talking with Barend Mons. Welcome. You just had a presentation here, and you were also part of the panel. One of the things that you focused on, was the European Open Science Cloud, because you are chairman of the Expert Group of the European Commission. So can you tell a little bit about the progress? What is the status of the European Science Cloud?

## 7.4. Critical success factors

The developments and expected impacts described above will not happen spontaneously. For these benefits to materialise a number of critical success factors (CSFs) must be in place. The following CSFs have been identified for EOSC:

- Researchers performing publicly funded research make relevant results available as openly as possible;
- Professional data stewards are available in research-performing organisations in Europe to help implement FAIR principles and support Open Science;

WORLD VIEW · 25 FEBRUARY 2020

## Invest 5% of research funds in ensuring data are reusable



It is irresponsible to support research but not data stewardship, says Barend Mons.

Barend Mons



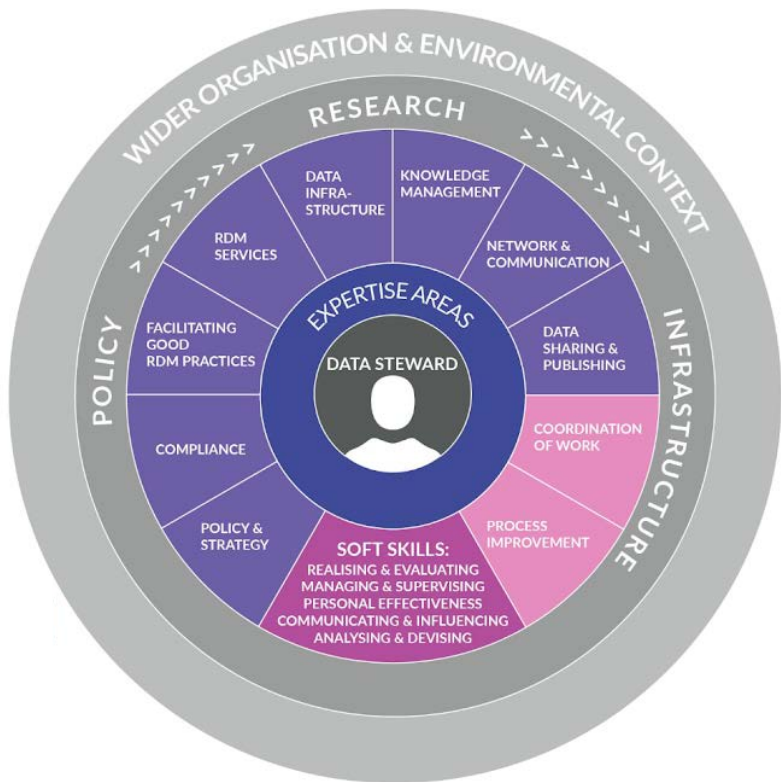
<http://www.doi.org/10.1038/d41586-020-00505-7>

to the European Commission. To be successful, the European Science Cloud needs a lot of experts to operate it. Barend Mons told us. Data stewards that have a lot of knowledge about managing and maintaining data. Experts who are well respected with a solid career path. Barend Mons also discussed several other findings of the Expert group, whose report will be published

# Chi è il «data steward\*» (1)

\***stewardship**: «pianificazione e gestione responsabile ed etica delle risorse».

Il «**data steward**» è una figura professionale competente nella pianificazione e gestione responsabile dei dati della ricerca durante tutto il loro ciclo di vita.



## Competenze

- Servizi di *research data management*, infrastrutture dati e di ricerca
- ✓ Gestione della conoscenza
- Conservazione e pubblicazione dei dati
- Conformità a leggi e normative istituzionali, nazionali ed internazionali in materia di gestione dei dati

## Abilità

- ✓ Organizzazione e coordinamento del lavoro
- ✓ Comunicazione e sensibilizzazione
- ✓ Gestione e miglioramento di processi
- ✓ Didattica e formazione

Mijke Jetten et al. (2021). Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship. Zenodo.  
<http://doi.org/10.5281/zenodo.4623713>

Valentina Pasquale – Istituto Italiano di Tecnologia

# Chi è il «data steward» (2)

## Profili professionali



### THE ADMINISTRATOR

- Establish good practices in compliance and data privacy
- Fast learner with a structured and analytical mindset
- Focus on execution and seek challenges in strategic development
- Implement solutions and educating end-users about them
- Passion for policy and IT security
- Positive attitude on cloud solutions
- Risk assessments while having disciplinary knowledge
- Team player with can-do attitude towards processes and operations



### THE ANALYST

- Ensure data quality
- Enthusiasm in cloud solutions
- Fast learner and innovative on building custom software and databases
- Good at multitasking
- Programming skills in statistical and data analysis
- Seek challenges, have positive attitude towards reporting



### THE DEVELOPER

- FAIR principles advisor and good at data planning and governance
- Focus on collaboration and knowledge sharing to raise business awareness
- Innovative thinker who develops procedures and guidelines
- Innovative thinking concerning master data management
- Passionate about process optimization via good project management
- Working in a team with compliance and data privacy experts trying to establish good practices



### THE AGENT OF CHANGE

- Agile mindset and enthusiasm
- Client and customer oriented, understanding both users and processes and operations
- Developing user friendly procedures and guidelines
- Educate users on ethics and the responsible conduct of research
- Focus on execution of policy and strategy awareness
- Passionate to implement solutions via project and change management

In generale, i «data steward»...

... possiedono **nozioni su come i dati vengono raccolti e gestiti** in un **dominio di ricerca specifico**

... conoscono gli **aspetti legali della gestione dei dati** (privacy, proprietà intellettuale) ed **etici**

... possibilmente hanno esperienza pregressa nella **ricerca, programmazione, gestione di database e infrastrutture di ricerca, sicurezza dei dati**

... hanno buone capacità **comunicative, didattiche e organizzative**

... **comprendono la psicologia dei ricercatori e parlano lo stesso linguaggio specifico**

... desiderano intraprendere un **percorso di carriera che non è né puramente scientifico né tecnico**



I "policy data stewards" (*amministratori e agenti di cambiamento*) sviluppano i regolamenti e le linee guida istituzionali (nazionali ed internazionali): i soggetti di riferimento sono i politici, gli enti finanziatori e il *management* di ricerca e universitario.

**AGENT OF CHANGE**  
Aligning researcher's data handling and data policies

I "research data stewards" (*analisti*) sono competenti nelle pratiche di data management specifiche per le discipline e lavorano a stretto contatto con i ricercatori.

  National Forum for Research  
  Data Management- Denmark

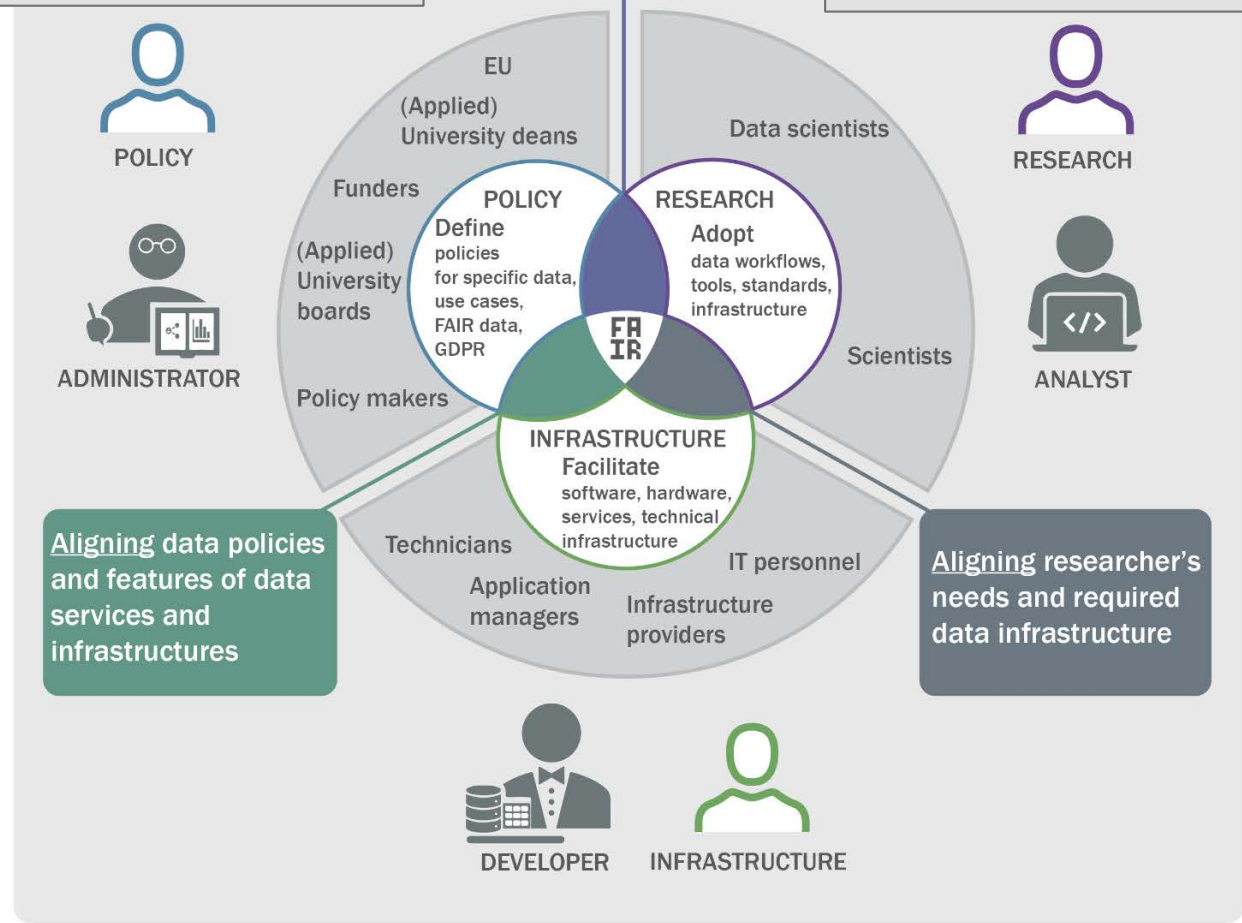


Four roles of a data steward

 ZonMw/ELIXIR project-  
 The Netherlands



Three roles of a data steward



Gli "infrastructure data stewards" (*sviluppatori*) hanno il ruolo di facilitatori e supervisionano la realizzazione dell'infrastruttura tecnica per i FAIR data.

Ruoli e competenze nei modelli olandesi e danesi

Modified from Frederike Schmitz. (2020). The roles of data stewards in the data stewardship landscape identified in Denmark and the Netherlands. Zenodo. <https://doi.org/10.5281/zenodo.4321265>

# Educazione dei «data steward» (1)

Martin Lillholm, Data Steward Education  
Copenhagen University, Challenges in  
professionalizing data stewardship BoF RDA  
P15

## Modelli di formazione ed educazione dei data steward

### Students with Bachelor degree

#### Directed corporate employment

- *One-year master programme*  
For students who fulfill requirements to basic programming skills, study skills, subject knowledge and academic language level.
- *Two year master programme*  
For students who do not meet the requirement for programming and study skills follow a *pre-master's year before the master's* and then continue with the one-year master.
- *Two-year candidatus*  
Traditional university candidatus combining theory, methods and internships. A dissertation in the area of DS is required.

### Students with a PhD or equivalent

#### Directed corporate or academic employment

For PhDs from any field at university faculties or part of research teams in industry. Educational programmes are a collaborative endeavor between faculties, library or knowledge centers, Centers for Information Security, Data Steward Community, Research coordinators, Project PI's, System Developers, Communication and Teaching teams, the Graduate School (for PhD training) & Human Research Ethics Committees.

- *Requirements:* Short courses, workshops, mentorships, online modules, summer school programmes & MOOCs, that immerse the student in the DS Community.

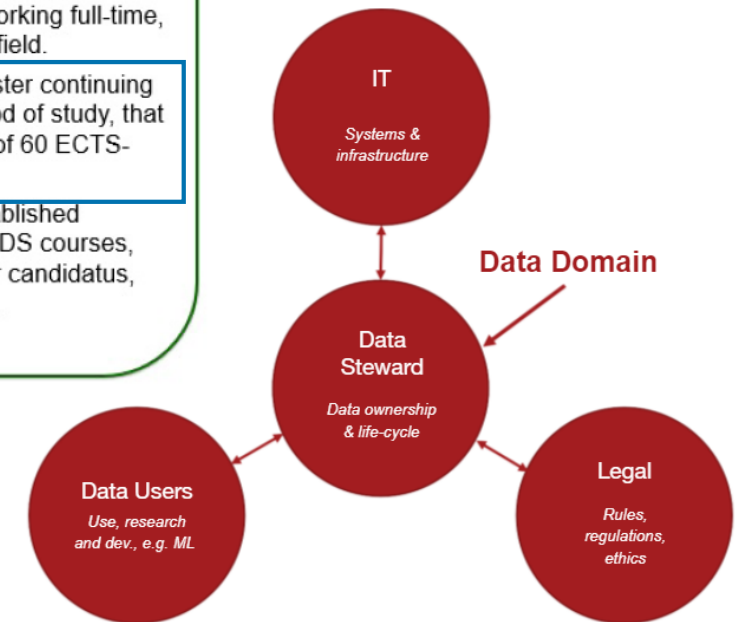
### Continuing and professional education

#### Directed corporate or academic employment

#### Flexible master programme

- For students who have professional experience and wish to improve their DS skills but keep working full-time, or already have a master's degree in any field.  
The programme is a part-time flexible master continuing education program with a prescribed period of study, that the student individually plans, and a total of 60 ECTS-points.
- The student combines elements from established programmes, thus it is a requirement that DS courses, such as the one year master and two year candidatus, are available through the Open University.

## Flessibilità, Integrazione, Formazione continua



Lorna Wildgaard et al. (2020). National Coordination of Data Steward Education in Denmark: Final report to the National Forum for Research Data Management. Zenodo.  
<https://doi.org/10.5281/zenodo.3609515>

# Educazione dei «data steward» (2)

Università de Montpellier

## Programme de la formation



### Comprendre l'environnement des données de la science

- Module 1 : Territoires, innovation et financement
- Module 2 : Recherche, gouvernance et données
- Module 3 : Données, innovation et droit

En savoir +

### Maîtriser les outils d'analyse des données de la science

- Module 1 : Extraire les données
- Module 2 : Gérer les données et les métadonnées
- Module 3 : Explorer et fouiller les données
- Module 4 : Sécuriser les données
- Module 5 : Protéger les données personnelles

En savoir +

### Gérer l'ouverture des données de la science

- Module 1 : Construire un plan de gestion de données (PGD)
- Module 2 : Stocker et archiver les données
- Module 3 : Diffuser et partager les données
- Module 4 : Valoriser et réutiliser les données

En savoir +



- ✓ How to find data more efficiently
- ✓ How to publish data Software and data carpentry
- ✓ How to use Github
- ✓ Introduction to theory and practice of visualization
- ✓ Introduction to machine learning
- ✓ Grounding in use of computational research infrastructures

- ✓ Principles and practices of Open Science
- ✓ FAIR data and using data ethically
- ✓ An introduction to research data management and the research data lifecycle
- ✓ How to prepare a data management plan
- ✓ Introduction to persistent identifiers and licensing
- ✓ A grounding in how to compile bibliographies

- Artificial Intelligence and Machine Learning
- Data Science and Applications
- Computational Science and Engineering

## Master degree in



DATA SCIENCE & SCIENTIFIC COMPUTING



- Data Management and Engineering

# Percorsi di carriera e modelli di sostenibilità

In alcuni paesi europei (Olanda, UK, Danimarca, Austria, etc.), la figura professionale del "data steward" è stata già introdotta all'interno dei Dipartimenti di ricerca e/o Facoltà e segue un **percorso di carriera separato dalle carriere scientifiche** e con un suo riconoscimento all'interno del **supporto alla ricerca**.

In generale, le Organizzazioni di Ricerca straniere che hanno già avviato questo percorso di introduzione dei "data steward" beneficiano di **fondi statali a tempo determinato per la fase di "start-up"**, legati all'implementazione di una pianificazione nazionale.

In quasi tutti i casi la **sostenibilità a lungo termine è lasciata alle Organizzazioni di Ricerca**, che dovranno implementare un modello di business che preveda l'assunzione a tempo indeterminato dei "data steward".

In **Italia**, attendiamo la pubblicazione del primo **Piano Nazionale Scienza Aperta** (anticipato in forma di abstract nel [Programma Nazionale per la Ricerca 2021-2027](#)), che contiene diversi pilastri, tra cui uno dedicato alla gestione dei dati della ricerca e che dovrebbe anticipare la messa a disposizione di risorse nazionali per supportare la Scienza Aperta.

29/10/2021 09:30-10:00  
Marta Teperek, Head of  
Research Data Services at  
TUDelft and  
4TU.ResearchData  
"Data Software roles of the  
future: Data stewardship and  
the Digital Competence Centre  
at TU Delft"

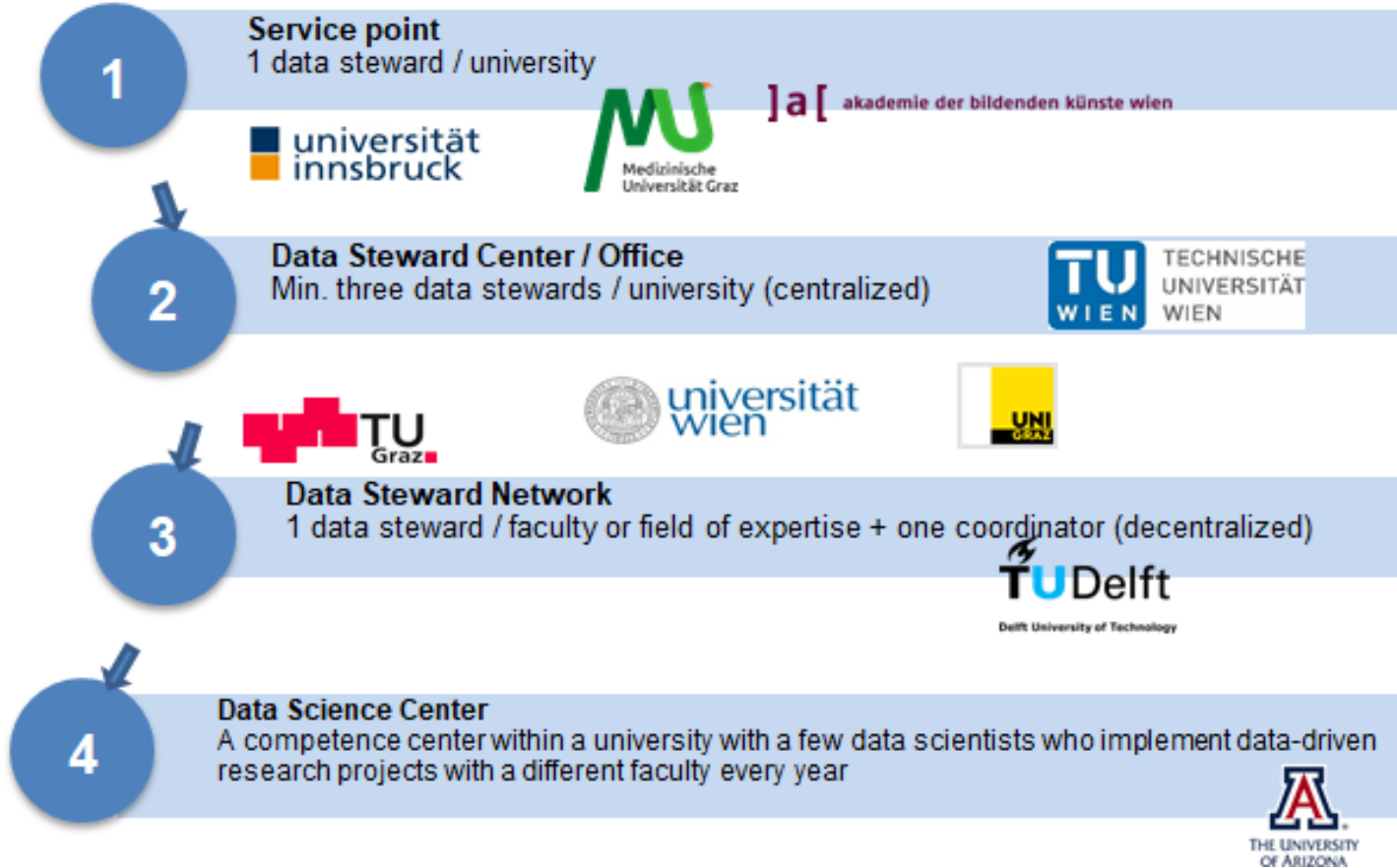
 TU Delft





Ricognizione dei modelli di "data stewardship" in diverse Università e Organizzazioni di Ricerca.

# Data Stewardship nelle università austriache



Slide credit: Ilire Hasani-Mavriqi, Sarah Stryeck (TU Graz)

Adattamento da: Buss, Mareike, Cragin, Melissa, Giglia, Elena, Hasani-Mavriqi, Ilire, Helling, Patrick, Mihai, Hannah, Pasquale, Valentina, Pawlowska, Maria M., Pignatari Drucker, Debora, Sesink, Laurents, Stryeck, Sarah, Walek, Anna, & Whyte, Angus. (2021, October 29). DSCC-IN National Chapters meeting slides, October 4-5 2021. Zenodo. <https://doi.org/10.5281/zenodo.5569096>

Valentina Pasquale – Istituto Italiano di Tecnologia

# Parola chiave: collaborazione

## Il ruolo delle *community of practice*

Rete internazionale dei *Data Stewardship Competence Centers*, focalizzata sulla convergenza e condivisione di iniziative territoriali nazionali.



RDA *Interest Group* per la professionalizzazione dei *data stewards*.



### Data Stewardship Competence Centers (DSCC)

<https://www.go-fair.org/implementation-networks/overview/dscc/>

**Professionalising Data Stewardship IG**  
Taxonomy:

- Posts
- Create Wiki index
- Events
- Repository
- Outputs
- Charter
- Plenaries
- Members

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

- TG1. a business case for data stewardship
- TG2. data stewardship terminology
- TG3. the integration of data stewardship across an organisation
- TG4. job profiles for data stewards
- TG5. training for data stewards
- TG6. career tracks for data stewards
- TG7. networking and knowledge exchange
- TG8. certification.



29/10/2021 10:30-11:15  
Emma Lazzeri, GARR,  
Coordinatrice del Centro di Competenza ICDI per l'Open Science, FAIR ed EOSC.  
"Il Centro di Competenza ICDI: formazione e supporto al servizio della scienza aperta e di EOSC"

International coordination



## Contatti

Valentina Pasquale

[valentina.pasquale@iit.it](mailto:valentina.pasquale@iit.it)

Research Data Management @IIT

[rdm@iit.it](mailto:rdm@iit.it)

# Grazie per l'attenzione!



**ISTITUTO ITALIANO DI TECNOLOGIA**

Sede Legale: Via Morego, 30 16163 Genova

Uffici di Roma: Via Guidubaldo del Monte, 54 00197 Roma

Tel. 010 2896

[www.iit.it](http://www.iit.it)

