

Progetto NFFA-DI - PNRR Missione 4, "Istruzione e Ricerca" - Componente 2, "Dalla ricerca all'impresa" - Linea di investimento 3.1, "Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione" - Azione 3.1.1, "Creazione di nuove IR o potenziamento di quelle esistenti che concorrono agli obiettivi di Eccellenza Scientifica di Horizon Europe e costituzione di reti" - CUP B53C22004310006.

D3.1 - REPORT ON THE FAIRNESS OF TOOLS WITHIN THE CONSORTIUM AND PLAN OF ACTION

BIM 5

OI 3.1 -THE FAIR DATA DIVISION (FAIR-
DD)

WP 3 - DIGITAL STRUCTURE OF NFFA-DI
AND OVERARCHING FAIR ECOSYSTEM
FOR DATA SYSTEM (OFED)



Piano Nazionale di Ripresa e Resilienza

Quadro riassuntivo rilasci documento

<i>Data</i>	<i>Stato documento</i>	<i>Realizzato da</i>	<i>Note</i>	<i>Supervisione</i>
30/11/2023	Draft	<i>Federica Bazzocchi, Mariarita de Luca</i>	First draft	
20/12/2023	Draft	<i>Federica Bazzocchi, Mariarita de Luca,</i>		Stefano Cozzini
21/12/2023	v.1.0.0	<i>Federica Bazzocchi, Mariarita de Luca, Ruggero Lot</i>		Stefano Cozzini

Sommario

1. INTRODUZIONE.....	4
2. MONITORAGGIO DELLO STATO DI FAIRNESS DEI NODI.....	5
2.1. PROCEDURA DI RACCOLTA DATI.....	5
2.2. DATI RACCOLTI.....	8
2.2.1. CNR-IOM.....	8
2.2.2. CNR-IFN@MI.....	16
2.2.3. CNR-IFN@TN.....	17
2.2.4. CNR-IMM@BO.....	18
2.2.5. CNR-IMM@CT.....	19
2.2.6. CNR-ISM.....	21
2.2.7. CNR-NANOTEC.....	22
2.2.8. CNR-SPIN.....	23
2.2.9. AREA SCIENCE PARK.....	24
2.2.10. POLIFAB-POLIMI.....	25
2.2.11. UNIMI.....	27
3. PIANO D'AZIONE.....	29
3.1. CLASSIFICAZIONE DELLE AZIONI NECESSARIE.....	29
3.1.1. GRUPPO A.....	30
3.1.2. GRUPPO B.....	31
3.1.3. GRUPPO C.....	31
3.1.4. GRUPPO D.....	31
3.2. COORDINAMENTO CENTRALIZZATO.....	31
3.3. CREAZIONE GRUPPI DI LAVORO.....	32
4. OBIETTIVI FAIRNESS E Considerazioni finali.....	32
5. BIBLIOGRAFIA.....	33
6. ANNEXES.....	33
6.1. FAIRNESS STATUS MONITORING.....	33
6.2. CENSIMENTO DI TECNICHE E STRUMENTI A CATALOGO NFFA-DI.....	35

1. INTRODUZIONE

Questo documento vuole dare una panoramica generale dello stato dell'arte del livello di FAIRness degli 11 nodi coinvolti nel progetto NFFA-DI. Facendo questo vuole delineare una serie di azioni specifiche e concrete per le diverse realtà sperimentali che possano portare al raggiungimento degli obiettivi del Work Package 3 (WP3) 'Digital Structure of NFFA-DI and FAIR Overarching Data System (OFDS)'.

Più nel dettaglio, questo documento è parte integrante della prima fase dell'attività A3.1 "The FAIR Data Division (FAIR-DD)" descritta qui di seguito

"Activity 3.1 contributes to the setup of the data infrastructure and the related data services of the OFED ecosystem that will be made available to the whole RI. This will be established within the ORFEO datacenter installed in Area Science Park by means of storage and hardware expansion as detailed in the budget.

All above mentioned software services will be developed, tested and validated using existing datacenter infrastructure setup. As soon as the hardware expansion will be completed, the set of services will be migrated in the new setup and will operate in "production mode".

Hardware definition and tender release:

In the first phase of the activity requirements of the general architecture and the hardware components of the expansion will be gathered and compared with the market best match alternative in terms of performance, flexibility and integration.

Hardware installation and configuration:

This phase integrates the expansion of hardware components on the ORFEO datacenter and relevant software configurations to run services, setting storage infrastructures and related software layers.

Data services developed and deployed:

Once data services will be developed, tested and validated, last phase consists of deployment of the first release of the OFED on the storage infrastructure."

Allo stesso tempo questo documento è propedeutico alle attività dalla A3.3 alla A3.12 relative all'implementazione dei principi FAIR-by-design nei singoli nodi.

All'interno della FDD nel deliverable D1.2 "Data Management Plan- First Release" del progetto NFFA-DI è stato introdotto il "FAIR Data Management Team", i cui partecipanti vengono indicati nella Tabella 1. Il team si è occupato raccogliere e analizzare i dati per redigere questo documento e stilare i vari piani d'azione. Nei mesi a venire darà inoltre supporto alle varie unità operative e monitorerà lo stato di avanzamento dei lavori durante lo svolgimento dell'intero progetto come descritto nella Sezione 3.2.

FAIR DATA MANAGEMENT TEAM (FDTM)	
Nome e Cognome	Istituto di appartenenza
Federica Bazzocchi	Area Science Park
Mariarita de Luca	Area Science Park
Ruggero Lot	Area Science Park
Dario De Angelis	CNR-IOM

Irene Modolo	CNR-IOM
------------------------------	---------

Tabella 1: composizione del FAIR Data Management Team.

2. MONITORAGGIO DELLO STATO DI FAIRNESS DEI NODI

2.1. PROCEDURA DI RACCOLTA DATI

Le informazioni necessarie per redigere questo documento sono state raccolte attraverso un questionario dedicato e raggiungibile previa registrazione sul sito del progetto NFFA-DI [1]. Ad ogni Unità Operativa (UO) del progetto NFFA-DI è stato chiesto di rispondere a 15 domande afferenti ai principi FAIR definiti in “ The FAIR Guiding Principles for scientific data management and stewardship” [2] declinati nella loro forma più estesa seguendo le indicazioni di “Explanation of the FAIR data principles”[3]. Per ogni principio FAIR, le domande sono state sdoppiate, laddove possibile, in modo da interrogare i ricercatori sia sulla fase di raccolta dei dati e metadati associati che sulla fase della loro pubblicazione. Così facendo i ricercatori sono stati invitati a rispondere a 24 domande di cui 7 afferenti alla *Findability (F)*, 4 all’ *Accessibility (A)*, 6 all’ *Interoperability (I)* e 7 alla *Reusability/Reproducibility (R)*. Per completezza il questionario è riportato nell’ Annex I.

Sono state raccolte le risposte di 23 dei 25 laboratori intervistabili così distribuiti rispetto agli 11 nodi:

UO	# Laboratori intervistati	# Laboratori intervistabili
AREA SCIENCE PARK	1	1
CNR-IFN@MI	1	1
CNR-IFN@TN	1	1
CNR-IMM@BO	1	1
CNR-IMM@CT	2	2
CNR-IOM	11	13
CNR-ISM	1	1
CNR-NANOTEC	1	1
PolifAB-POLIMI	1	1
UNIMI	2	2
CNR-SPIN	1	1
Totale	23	25

Tabella 2: distribuzione riassuntiva delle risposte raccolte e delle risposte attese.

Per comprendere la complessità dell’infrastruttura di ricerca prevista, nella tabella che segue riportiamo tutti i laboratori che partecipano al progetto NFFA-DI, la loro UO di riferimento e i rispettivi Responsabili di Laboratorio. Abbiamo evidenziato in rosso le caselle dei laboratori per i quali non è stato possibile raccogliere le risposte. Per i singoli laboratori, laddove è stato possibile, vengono indicate le tecniche, lo strumento sperimentale e le tecniche ancillari legate all’attrezzatura aggiuntiva. Tale suddivisione è allineata all’organizzazione a più livelli del

catalogo di NFFA-DI che sarà disponibile agli utenti che accederanno al Portale Unico. In accordo col WP2 si sono definiti

- *tecnica (technique)* la tecnica sperimentale per cui un utente fa richiesta di accesso all'infrastruttura;
- *strumento (instrument)* la strumentazione sperimentale che permette lo svolgimento della tecnica;
- *attrezzatura ancillare (ancillary equipment)* le tecniche aggiuntive che vengono effettuate a supporto o completamento della principale, che possono avere strumentazione propria o no, ma che non rappresentano la motivazione per cui un utente sceglie di accedere all'infrastruttura.

Nodo	Laboratorio	Tecniche	Strumento	Ancillary Equipment	Responsabile di Laboratorio
CNR-IOM	APE HE beamline	XAS, XMCD		LEED	Piero Torelli
	APE Cluster	PES, MOKE		MBE, RHEED	Cinzia Cepek
	APE LE	PES		LEED	Ivana Vobornik
	INSPECT	PES		LEED	Cinzia Cepek
	STRAS - LEEM-PEEM	LEEM-PEEM		LEEM, PEEM	Alessandro Sala
	STRAS - STM-VT	STM		PES, LEED	Mirco Panighel
	SPRINT - RAMAN	RAMAN			Andrea Giugni
	SPRINT - TRANSIENT GRATING	TRANSIENT GRATING			Riccardo Cucini
	SPRINT - PHOTOLUMINESCENCE	PHOTOLUMINESCENCE			Riccardo Cucini
	SPRINT - Mott spectroscopy	TR-ARPES		PES, ARPES, SPIN-POL	Gian Marco Pierantozzi
	HMMBE	HMMBE		MBE, RHEED	Giorgio Biasiol
	MBE/XPS	MBE/XPS		MBE	Silvia Rubini
CNR-IFN@MI	UDYNI				Michele Devetta
CNR-IFN@TN	CleanRoom CRD @ FBK	LITHOGRAPHY & PATTERNING- RIE			Lorenza Ferrario
CNR-IMM@BO					Stefano Zampolli
CNR-IMM@CT	Experimental	ATOM PROBE TOMOGRAPHY, STRUCTURAL AND ANALYTICAL IMAGING			Antonino La Magna
	Theory - Simulation				Antonio Mio
CNR-ISM	EuroFEL Support Lab				Stefano Turchini
CNR-NANOTEC	Clean Room E (nanofab)	SEM		FIB	Gianluca Balestra
CNR-SPIN					Francesco M. Taurino
AREA SCIENCE PARK	LAME	TEM, SEM		FIB-SEM	Federica Bazzocchi
POLIMI	PolifAB				Matteo Cantoni

UniMI	Exp MiMAG	SEM-PA, XPS, RAMAN			Andrea Giugni
	Theory MLMS				Giovanni Onida

Tabella 3: dettaglio dei laboratori inseriti nel progetto NFFA-DI.

Nel loro insieme le Tabelle 2 e 3 sintetizzano gli indicatori misurabili associati a questo deliverable e corrispondenti agli obiettivi intermedi legati all'attività A3.1.

La procedura introdotta ha permesso di analizzare le risposte dei singoli laboratori in modo da trarre delle conclusioni specifiche nonché aggregate.

Per ottenere delle indicazioni quantitative in modo da poter definire degli obiettivi concreti si è introdotto il seguente criterio di assegnazione dei punteggi:

- ad ogni domanda poteva essere assegnato un valore da 0 a 1;
- l'intervallo unitario è stato suddiviso per il numero di domande secondarie contenute nelle domande principali (ad esempio se si avevano 2 domande secondarie, ogni domanda secondaria valeva un massimo di 0,5 punti);
- ogni intervallo delle domande secondarie è stato ulteriormente suddiviso qualora ci fossero due tipologie di tecniche o due strumentazioni diverse o una serie di casi specifici elencati e i ricercatori avessero specificato i protocolli applicati nei singoli casi (ad esempio 2 tecniche diverse per una domanda con 2 domande secondarie faceva sì che ogni domanda secondaria valesse al massimo 0,25 punti);
- a questo punto, la risposta alle domande 'atomiche' definite in questo modo assumeva un valore booleano;
- in alcuni casi i ricercatori sono stati interrogati su laboratori ancora non operativi: in questi casi le risposte che riportavano l'intenzione a implementare una determinata prassi di lavoro o procedura supportata da una spiegazione dettagliata sono state valutate a punteggio pieno.
- in svariati casi i ricercatori in riferimento a un laboratorio già esistente hanno espresso l'intenzione di implementare determinate procedure o best practice. Ai fini del punteggio queste intenzioni sono state valutate nulle, ma se ne è tenuto conto nella definizione del piano d'azione.

Nel questionario i ricercatori erano invitati a rispondere il più possibile in maniera completa e dettagliata.

Il modello di assegnazione del punteggio descritto precedentemente ha permesso di rappresentare lo stato di FAIRness di ciascun laboratorio con un grafico radar che da' una percezione visiva dello stato complessivo di FAIRness dei singoli laboratori. Nel caso perfetto i punteggi si distribuiscono sulla circonferenza più esterna in modo che l'area sottesa dall'unione dei punti sia massima. Minore l'area, inferiore la FAIRness. Chiaramente questo tipo di rappresentazione sfavorisce i laboratori che al momento soddisfano i principi FAIR a singhiozzo, ad esempio solo per i dati pubblicati e non per quelli raccolti o viceversa. Tuttavia, ha un impatto visivo estremamente intuitivo.

Le risposte sono state riassunte riportando nell'analisi che segue le informazioni principali corrispondenti ai quattro principi cardine dei principi FAIR.

Per ogni laboratorio che ha risposto al questionario le risposte sono state riassunte riportando nell'analisi le informazioni principali corrispondenti ai quattro principi cardine dei principi FAIR (Findability, Accessibility, Interoperability, Reusability/Reproducibility). Per ciascun principio

cardine viene indicato anche lo score medio ottenuto assegnando un punteggio alle 24 risposte in base ai criteri e alla classificazione precedentemente esposti.

Viene inoltre indicato lo score medio globale. Come vedremo nella sezione 3.1 quest'ultimo ha permesso la clusterizzazione dei laboratori in laboratori che condividono esigenze simili.

2.2. DATI RACCOLTI

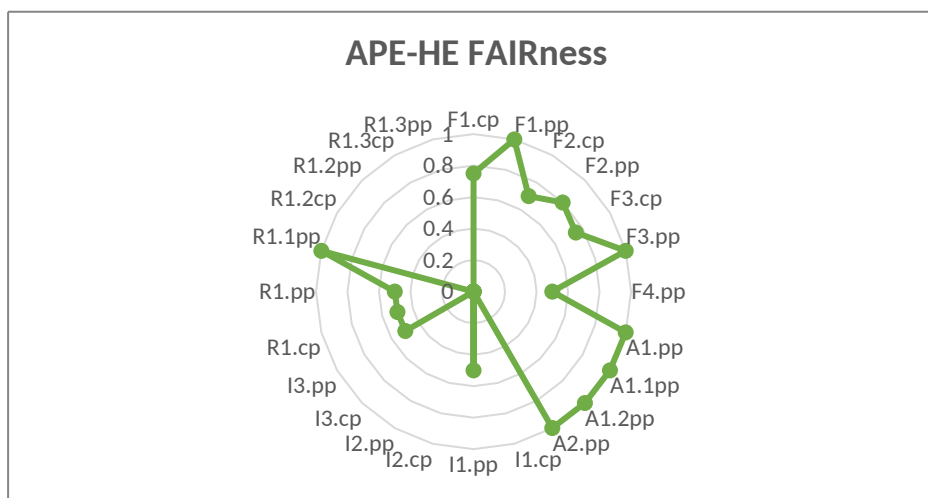
2.2.1. CNR-IOM

Il CNR-IOM partecipa al progetto con 13 laboratori censiti di cui 11 hanno risposto al questionario.

Per esigenze operative alcuni laboratori hanno dato risposte cumulative e vengono pertanto analizzati insieme in quanto segue.

APE-HE

Principio	Situazione Globale	Score
F	Uso di un identificativo univoco per la raccolta dei dati che usa quasi sempre la stessa convenzione. I raw data e i metadati raccolti automaticamente dal software sono pubblicati su repository con DOI ma altri metadati non sono pubblicati seppur salvati su server locali correttamente collegati ai rispettivi dati, ma senza riferimento ai DOI. I metadati pubblicati coi raw data e quelli richiesti in fase di sottomissione su Zenodo sono comunque generici e non esplicativi dei dati raccolti.	0,81
A	I dati collegati a una pubblicazione e i corrispettivi metadati acquisiti automaticamente dal software sono pubblicati su Zenodo. L'accesso è regolamentato dai protocolli della piattaforma.	1
I	Non si usa uno standard per la raccolta dei metadati se non quello definito dal software di acquisizione. I dati pubblicati portano il riferimento alle pubblicazioni collegate.	0,17
R	Alcuni metadati sono salvati su notebook elettronici ma non vengono usati standard della comunità scientifica di riferimento. In fase di pubblicazione dei dati e dei metadati (acquisiti dal software) viene chiaramente indicata la licenza.	0,29
Media	Maggiore carenza: relativa alla fase di pubblicazione dei metadati	0,52

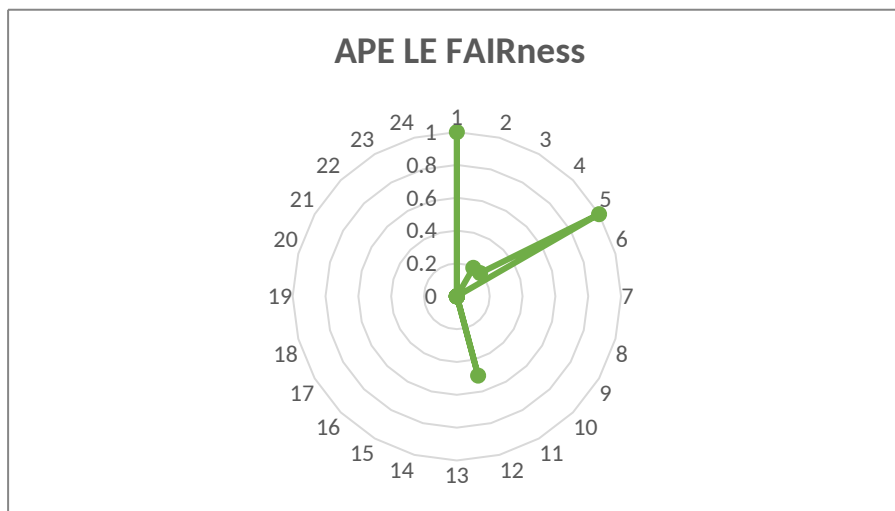


APE Cluster

Nessuna risposta ricevuta.

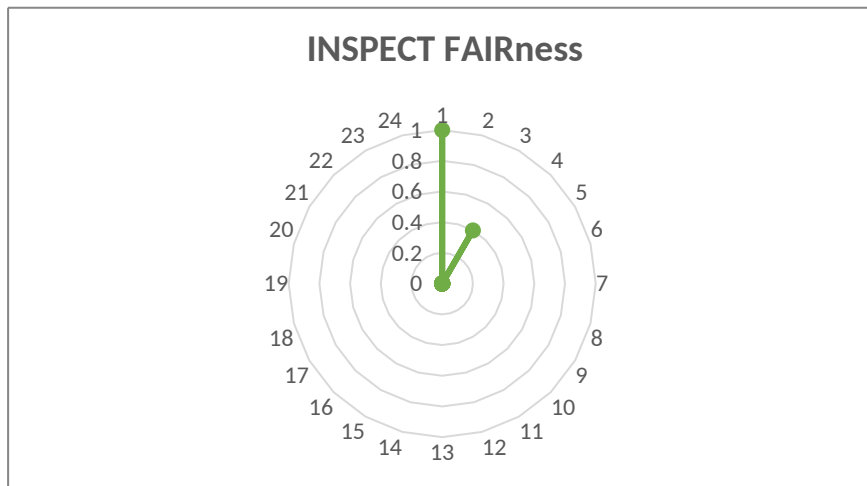
APE LE

Principio	Situazione Globale	Score
F	Viene usata una nomenclatura standard interna al laboratorio, ma dati e metadati sono solo parzialmente pubblicati. I metadati vengono raccolti in parte automaticamente dal software di acquisizione e in parte in un quaderno di laboratorio cartaceo. La corrispondenza fra dati e metadati è ricostruibile solo all'interno del laboratorio.	0,34
A	Dati e metadati non sono pubblicati. I dati e i metadati acquisiti automaticamente sono accessibili su richiesta attraverso servizi cloud esterni, mentre per i metadati salvati manualmente bisogna aver accesso al laboratorio.	0,00
I	Fatta eccezione per i metadati salvati automaticamente dalla beamline, i metadati sono in formato cartaceo e non pubblicati, quindi non interoperabili.	0,08
R	Si trova traccia dell'origine dei dati nella descrizione dei workflow negli articoli pubblicati, ma non essendoci metadati pubblicati la riproducibilità è totalmente compromessa.	0,00
Media	Carenza: sia nella fase di raccolta che pubblicazione dei dati e metadati	0,12



INSPECT

Principio	Situazione Globale	Score
F	Viene usata una nomenclatura standard interna al laboratorio, ma dati e metadati non vengono pubblicati. I metadati vengono raccolti ma in un quaderno di laboratorio cartaceo.	0,20
A	Dati e metadati non sono pubblicati. I dati sono accessibili su richiesta attraverso servizi cloud esterni, mentre per i metadati bisogna aver accesso al laboratorio.	0,00
I	Fatta eccezione per alcuni metadati salvati automaticamente, i metadati sono in formato cartaceo, quindi non interoperabili.	0,00
R	Si trova traccia dell'origine dei dati nella descrizione dei workflow negli articoli pubblicati, ma non essendoci metadati pubblicati la riproducibilità è totalmente compromessa	0,00
Media	Carenza : sia nella fase di raccolta che pubblicazione dei dati e metadati	0,0 6

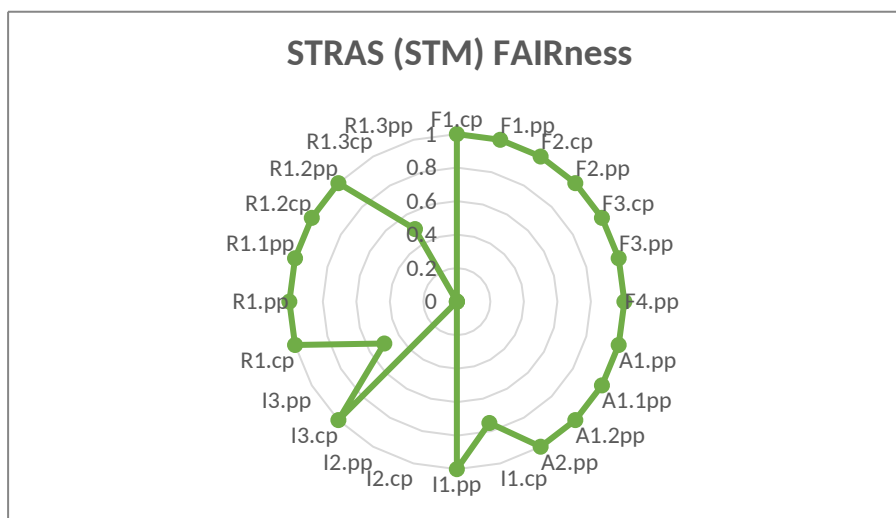


STRAS (LEEM-PEEM)

Nessuna risposta ricevuta perché il set-up sperimentale non esiste ancora.

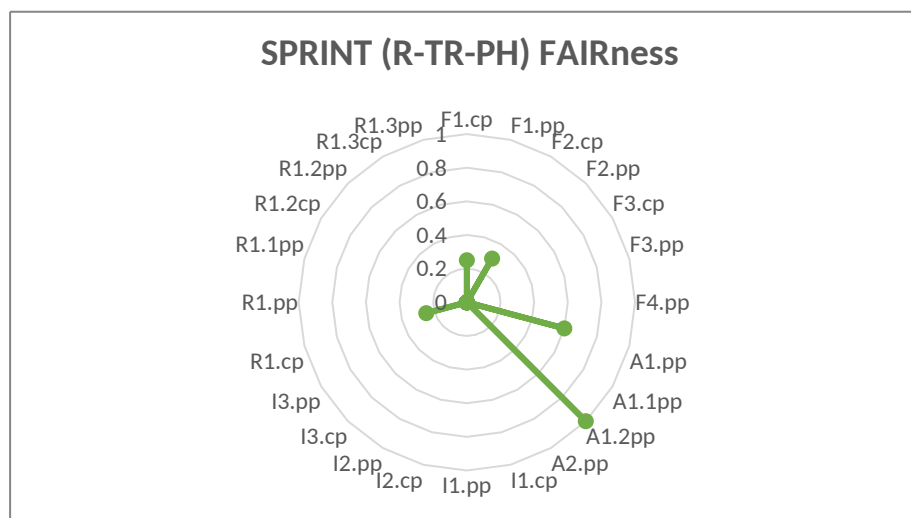
STRAS (STM-LT, STM-VT)

Principio	Situazione Globale	Score
F	Viene utilizzato un identificativo univoco per dati e metadati sia raccolti che pubblicati. I metadati sono pubblicati insieme ai corrispettivi dati oppure con riferimento agli stessi. Sono completi.	1,00
A	Dati e metadati sono pubblicati su Zenodo con licenze appropriate. L'accesso secondo i principi FAIR è garantito dal repository.	1,00
I	L'acquisizione avviene principalmente automaticamente ed è in fase di sviluppo uno schema di metadati basato su Nexus.	0,54
R	I metadati contengono chiari riferimenti alla strumentazione e agli strumenti. I dati e i metadati sono disponibili su Zenodo e lo schema è in fase di sviluppo	0,79
Media	Unica carenza: conclusione della definizione del metadata schema	0,82



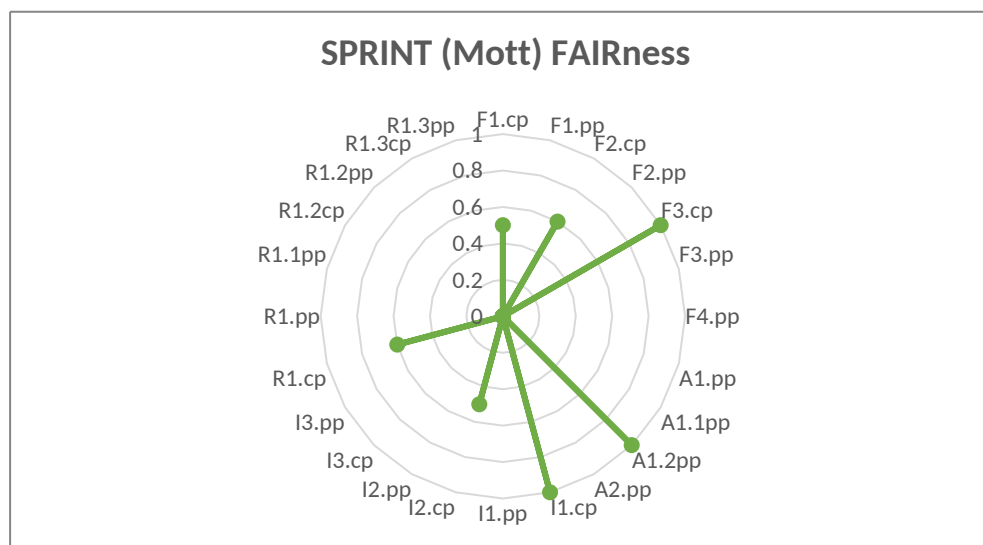
SPRINT (Raman-R-,Transient grating-TR-, Photoluminescence-PH-)

Principio	Situazione Globale	Score
F	Dati e metadati vengono raccolti senza una nomenclatura convenzionale e non sono pubblicati con un DOI. Solo in alcuni casi i metadati sono salvati in un logbook, ma non sono pubblicati.	0,08
A	I dati sono depositati sul datashare di NFFA-Eu e accessibili previa richiesta. Si pianifica di utilizzare in futuro delle repository ufficiali.	0,4
I	Non si usa uno standard nella raccolta dei metadati che però sono in un semplice formato testuale.	0
R	Gli unici metadati raccolti sono nel logbook e non c'è riferimento ai dati o alle pubblicazioni. Inoltre, in un caso i dati sono leggibili solo da software proprietario. Non vengono usati schemi di metadati standard.	0,04
Media	Carenza: sia nella fase di raccolta che di pubblicazione dei dati e metadati.	0,10



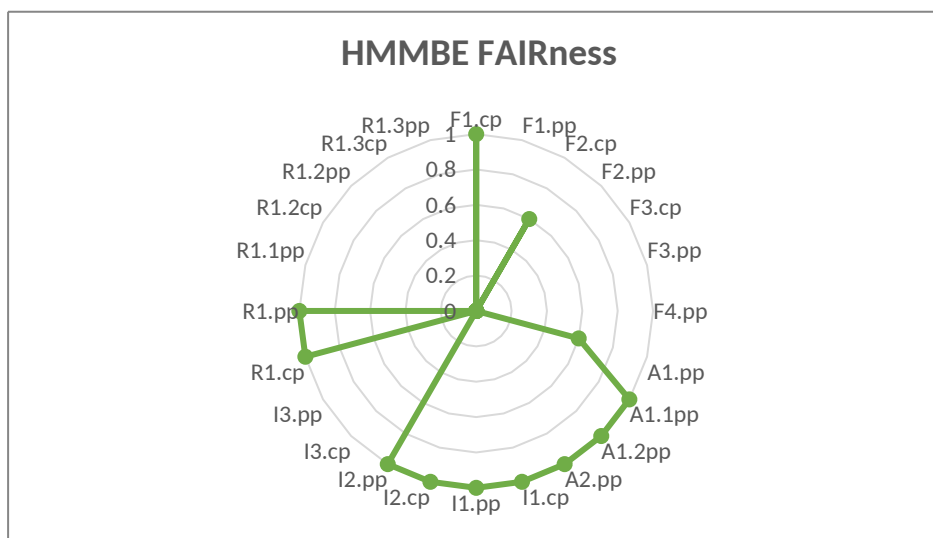
SPRINT (Mott)

Principio	Situazione Globale	Score
F	Dati e metadati raccolti con una nomenclatura convenzionale solo per una delle due tipologie di esperimento, ma non pubblicati. Metadati raccolti alcuni su carta e poi salvati coi dati.	0,30
A	Dati e metadati salvati solo sul datashare di NFFA-EU. Su richiesta viene dato accesso ad utenti esterni ma devono sapere che i dati esistono.	0,25
I	Si usa parzialmente uno standard nella raccolta dei metadati, che li rende interpretabili solo ad esperti del settore. Non si inseriscono riferimenti ad altri datasets rilevanti.	0,25
R	I metadati sono acquisiti ma non tutti i parametri rilevanti vengono salvati. Non si tiene traccia dell'origine dei dati e non essendo pubblicati non hanno licenza.	0,09
Media	Carenza: sia fase raccolta che pubblicazione	0,22



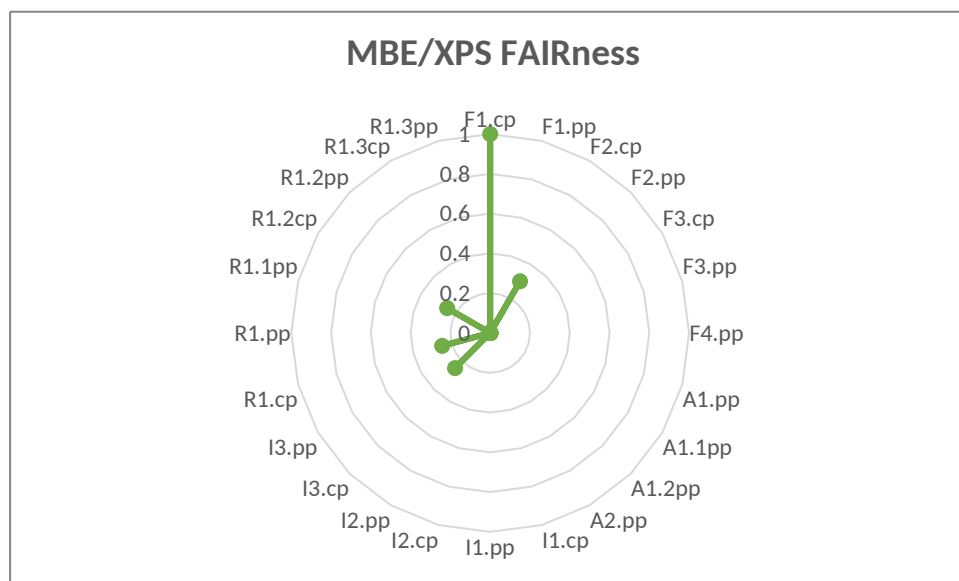
HMMBE

Principio	Situazione Globale	Score
F	Dati e metadati raccolti con una nomenclatura convenzionale, ma non pubblicati. Utilizzato un quaderno di laboratorio elettronico e aperto. Metadati ricchi di informazioni ma mancano di un identificativo univoco di collegamento coi datasets.	0,23
A	Dati e metadati salvati solo sul datashare di NFFA-EU. Su richiesta viene dato accesso ad utenti esterni ma devono sapere che i dati esistono.	0,90
I	Si usa uno standard nella raccolta dei metadati. Non si inseriscono riferimenti ad altri datasets rilevanti.	0,67
R	I metadati sono arricchiti di tutti i parametri rilevanti. Non si tiene traccia dell'origine dei dati e non essendo pubblicati non hanno licenza. Gli standard utilizzati non sono però standard riconosciuti dalla comunità di riferimento.	0,29
Media	Maggiore carenza: fase di pubblicazione di dati e metadati.	0,47



MBE/XPS

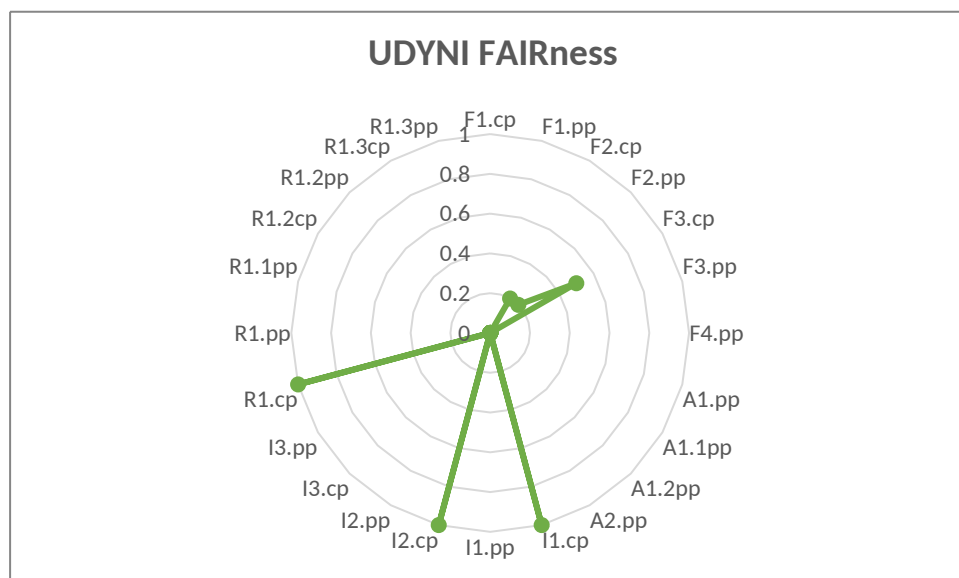
Principio	Situazione Globale	Score
F	Dati e metadati raccolti con una nomenclatura convenzionale ma non sono pubblicati con un DOI. Metadati in parte raccolti su notebook elettronici in parte su excell. Solo in alcuni casi i metadati sono salvati in un logbook, ma non sono pubblicati.	0,19
A	Dati salvati solo su server locali non accessibili ad utenti esterni.	0
I	Non si usa uno standard nella raccolta dei metadati e solo in alcuni casi viene fatto riferimento ai dati corrispondenti nel logbook elettronico.	0,04
R	Gli unici metadati raccolti sono nel logbook e solo se ritenuto importante c'è un riferimento ai dati o al processo. Inoltre in un caso i dati sono leggibili solo da software proprietario. Non vengono usati schemi di metadati standard. I grafici finali pubblicati mancano di provenance e la licenza non è esplicitata.	0,07
Media	Carenza: sia nella fase di raccolta che di pubblicazione di dati e metadati.	0,09



2.2.2. CNR-IFN@MI

UDYNI

Principio	Situazione Globale	Score
F	Il laboratorio è nuovo e non ancora operativo, per cui non è stata ancora definita una nomenclatura. Si pianifica la raccolta di metadati in maniera automatica e l'utilizzo di un quaderno di laboratorio elettronico. Si sta pensando di utilizzare formati che permettano di avere dati e metadati insieme, come HDF5.	0,13
A	Non è stato ancora definito nessun protocollo di accesso ai dati.	0,00
I	E' pianificato di acquisire dati e metadati attraverso delle pipeline automatiche in modo da salvarli insieme in file HDF5. I metadati verranno salvati seguendo degli standard. Al momento non è pianificato di pubblicare i dati con riferimenti univoci ad altri dati e metadati associati.	0,33
R	I metadati saranno arricchiti di tutti i parametri rilevanti. Non è stato ancora deciso di come tener traccia dell'origine dei dati, di quali licenze utilizzare e se usare dei metadata schema riconosciuti dalla comunità di riferimento.	0,14
Media	Carenza: sia nella fase di raccolta che di pubblicazione di dati e metadati.	0,17

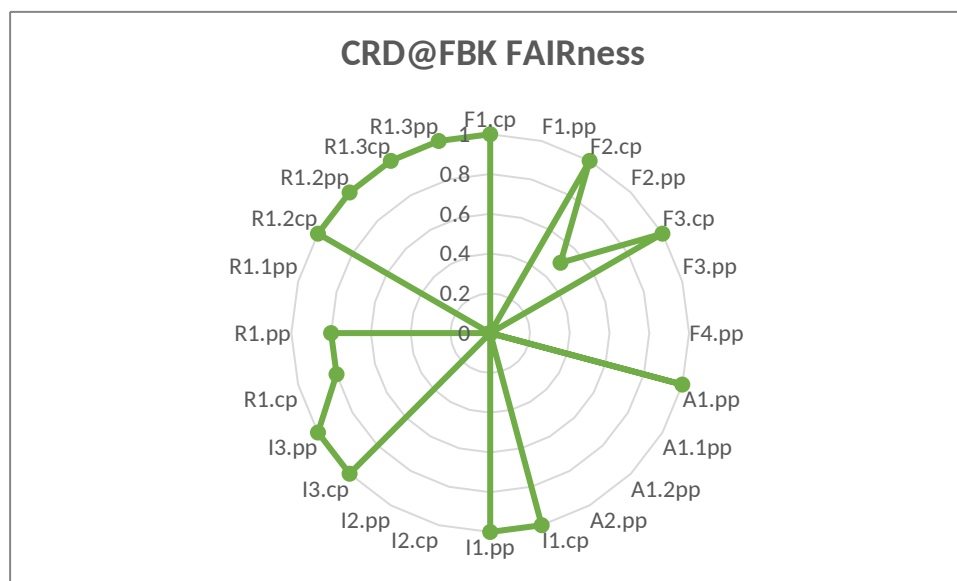


2.2.3. CNR-IFN@TN

Cleanroom CRD located at FBK

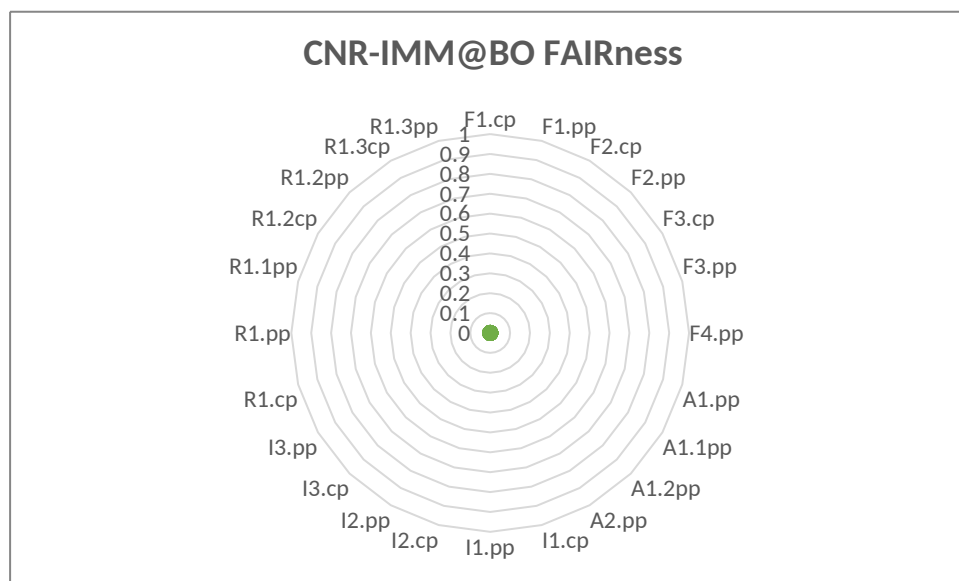
Principio	Situazione Globale	Score
F	Tutti i dati e i metadati sono raccolti dalla fabLIMS anche se poi non vengono pubblicati con un PID. Il collegamento fra dati raccolti e metadati avviene solo a livello di fase di raccolta. La fabLIMS agisce da quaderno elettronico di laboratorio e tiene traccia di tutti i metadati associati al campione e agli strumenti. I metadati sono però interpretabili chiaramente solo con informazioni aggiuntive.	0,50
A	E' possibile accedere ai dati tramite API solo per gli utenti registrati alla fabLIMS. Il software è chiuso.	0,25
I	Imetadati sono salvati attraverso pipeline automatizzate su un DB relazionale e possono essere esportati in formati standard. Non sono definite ontologie, ma sono inseriti riferimenti ad altri datasets rilevanti.	0,67
R	I metadati sono arricchiti di tutti i parametri rilevanti, in particolare è ben tracciata l'origine dei dati. I metadati seguono degli standard riconosciuti dalla comunità di riferimento, ma non sono definite le licenze per i dati pubblicati.	0,80

Media	Maggiore carenza: fase pubblicazione dati e metadati	0,59



2.2.4. CNR-IMM@BO

Principio	Situazione Globale	Score
F	Dati e metadati non sono pubblicati, i metadati non vengono raccolti ma c'è massima apertura ad ottenere supporto dal gruppo di Data Management di NFFA-DI. Mancanza anche di un server locale per la raccolta di dati e metadati.	0,00
A	I dati e metadati non sono pubblicati per cui non sono accessibili.	0,00
I	I metadati non sono raccolti per cui non è soddisfatto nessun criterio relativo alla interoperabilità.	0,00
R	I metadati non sono raccolti per cui non è soddisfatto nessun criterio relativo alla riproducibilità e riusabilità dei dati.	0,00
Media	Carenza: sia nella fase di raccolta che di pubblicazione di dati e metadati.	0,00

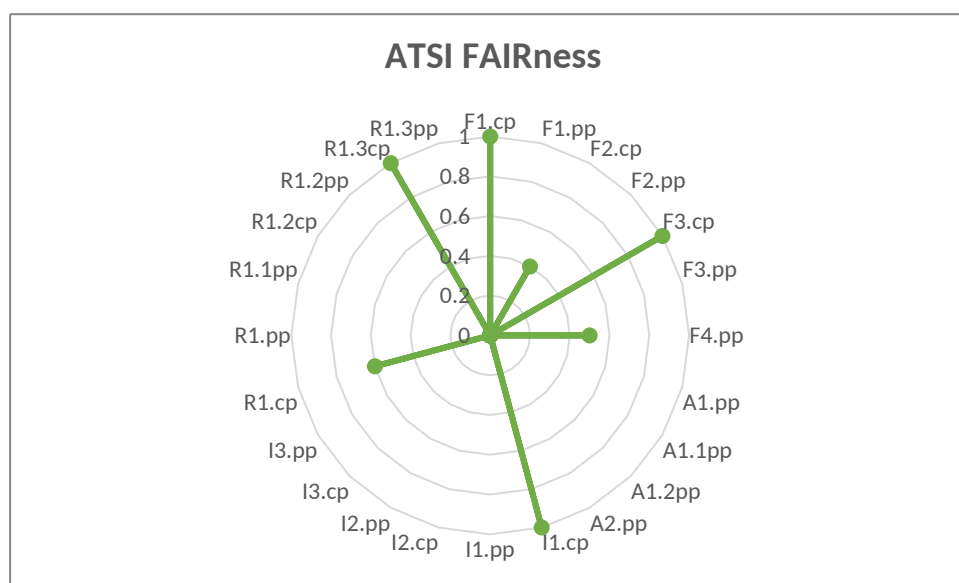


2.2.5. CNR-IMM@CT

Atom probe tomography, structural and analytical imaging (nel grafico seguente ATSI)

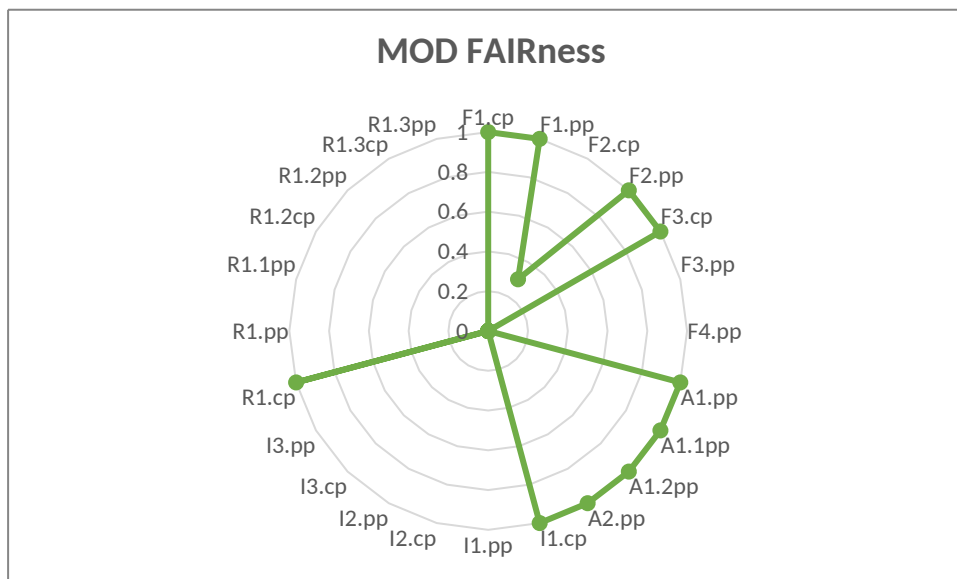
Principio	Situazione Globale	Score
F	Metadati acquisiti automaticamente dalla pipeline per cui seguono una nomenclatura automatica, ma non sono pubblicati. Vengono raccolti solo i metadati relativi alla fase di acquisizione dell'immagine, non quelli del sample. Si fa uso di un quaderno di laboratorio elettronico ma non è aperto. Note aggiuntive sono collegate in maniera univoca al campione e ai dati, ma solo all'interno del laboratorio. Metadati inclusi nei dataset in formati proprietari ma hanno realizzato delle librerie open per estrarli.	0,41
A	I metadati non sono accessibili al di fuori del laboratorio. Per i dati non sono fornite informazioni su come avervi accesso.	0,00
I	I metadati sono raccolti dal software di acquisizione e usano pertanto le logiche del produttore e non ontologie standard. Non vengono inseriti collegamenti ad altri dataset rilevanti. Il campione è fornito da altro produttore senza alcun metadato.	0,17
R	I metadati raccolti contengono le informazioni principali anche se non tutte. I metadati sono raccolti in maniera analoga a quanto fa la comunità di riferimento anche se non sono usate delle ontologie.	0,23

Media	Maggiore Carenza: fase di pubblicazione di dati e metadati.	0,23
--------------	---	------



Modelling (nel grafico seguente MOD)

Principio	Situazione Globale	Score
F	Si tratta di laboratorio che effettua simulazioni. La nomenclatura è standard all'interno del protocollo e dati e progetti sono caricati su GITHUB e referenziati direttamente dalle pubblicazioni. Non si fa uso di un quaderno di laboratorio elettronico. Localmente i dati sono in un folder dentro il folder dei metadati. I metadati non vengono pubblicati.	0,61
A	L'utilizzo di GITHUB permette rende accessibili i dati e i progetti secondo protocollo e policy standard.	1,00
I	Si tratta di simulazioni per cui per definizione i metadati sono raccolti in maniera automatica. Non si fa però uso di ontologie standard. Non vengono inseriti collegamenti ad altri dataset rilevanti.	0,17
R	Il riutilizzo dei dati è garantito facendo uso di GITHUB. Non vengono dichiarate le licenze, né utilizzati standard riconosciuti dalla comunità.	0,17
Media	Maggiore carenza: fase pubblicazione metadati	0,43

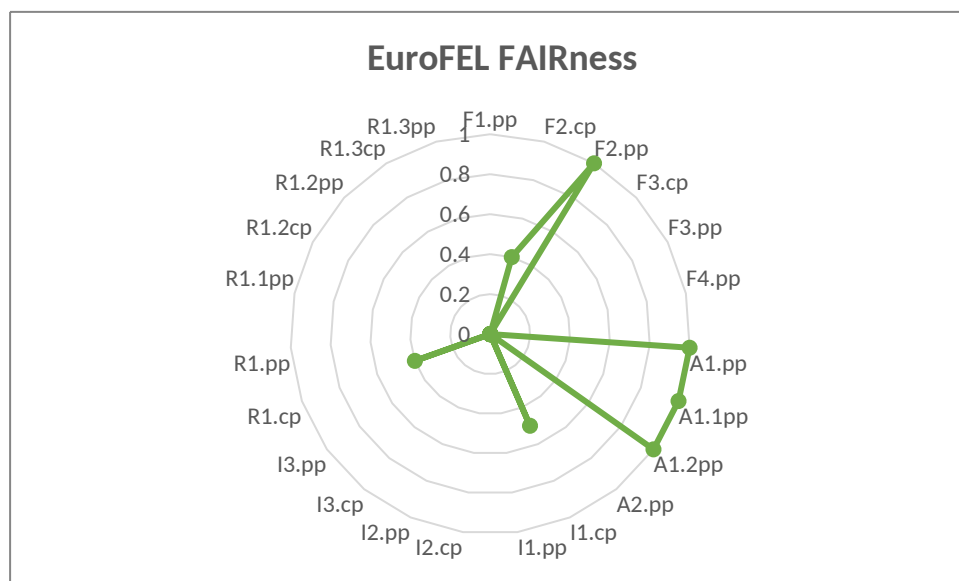


2.2.6. CNR-ISM

EuroFEL Support Lab

Principio	Situazione Globale	Score
F	Dati e metadati raccolti utilizzando una nomenclatura standard ma non tutto è pubblicato con un PID. Come quaderno di laboratorio elettronico viene usato un documento Google dove vengono salvati alcuni metadati. I dati vengono pubblicati corredati da alcune informazioni essenziali. Non si identificano ulteriori metadati.	0,34
A	I dati sono archiviati e condivisi su un cloud CNR tramite ownCloud, basato su tecnologia open source. Utenti esterni possono avere accesso all'ownCloud.	0,75
I	Fatta eccezione per le informazioni scritte sul documento Google tutti i (meta)dati vengono raccolti automaticamente. Non sono definite ontologie e inseriti riferimenti ad altri datasets rilevanti.	0,08
R	La struttura dei dati è relativamente semplice e informazioni aggiuntive sono contenute nel documento Google. I nomi delle variabili le autodefiniscono. I metadati non ricostruiscono l'origine dei dati e non vengono forniti utilizzando schemi standard.	0,06

Media	Maggiore carenza: fase pubblicazione di dati e metadati.	0,26

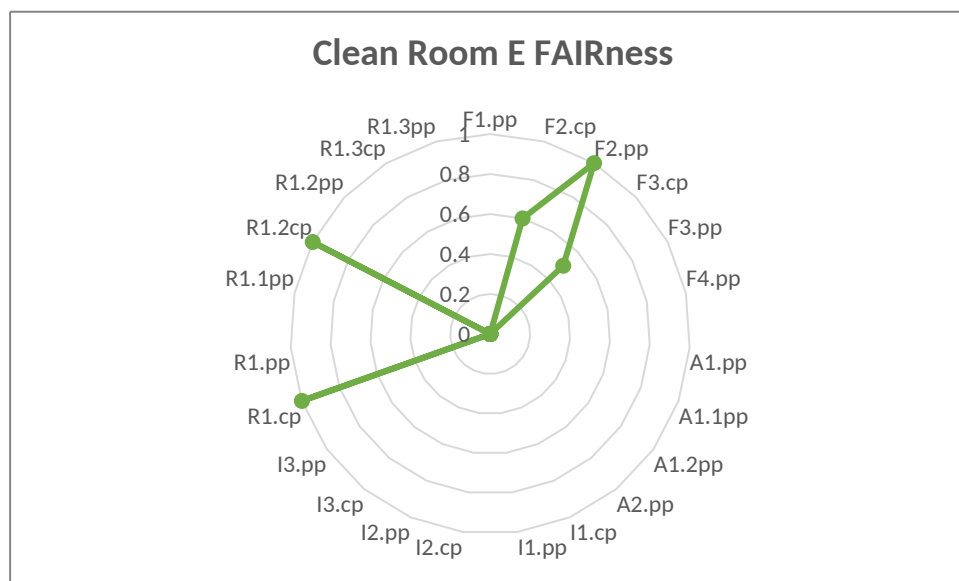


2.2.7. CNR-NANOTEC

Clean Room E (nanofab)

Principio	Situazione Globale	Score
F	Dati e metadati raccolti utilizzando una nomenclatura standard ma al momento niente è pubblicato con un PID. Parte dei metadati sono raccolti direttamente dallo strumento in formato proprietario. Altri metadati sono salvati su un foglio di calcolo Google ed è pianificato per un futuro l'introduzione di un quaderno di laboratorio elettronico. Vengono salvati i metadati dei campioni e altre informazioni rilevanti. Stanno testando un approccio di rete per connettere dati sul server e metadati della strumentazione.	0,44
A	I dati e i metadati non sono al momento pubblicati e sono solo salvati su server locali. A tali server può essere dato accesso ad utenti esterni tramite protocolli standard, dopo richiesta esplicita.	0,00
I	I metadati non vengono raccolti da una pipeline automatica e non sono stati definiti schemi di metadati appositi.	0,00

R	I metadati sono raccolti in maniera esaustiva e viene ricostruita in parte l'origine dei dati. Tuttavia non vengono strutturati utilizzando schemi standard e non essendo pubblicati impediscono di fatto la riutilizzabilità e la riproducibilità.	0,29
Media	Maggiore carenza: fase pubblicazione di dati e metadati.	0,21

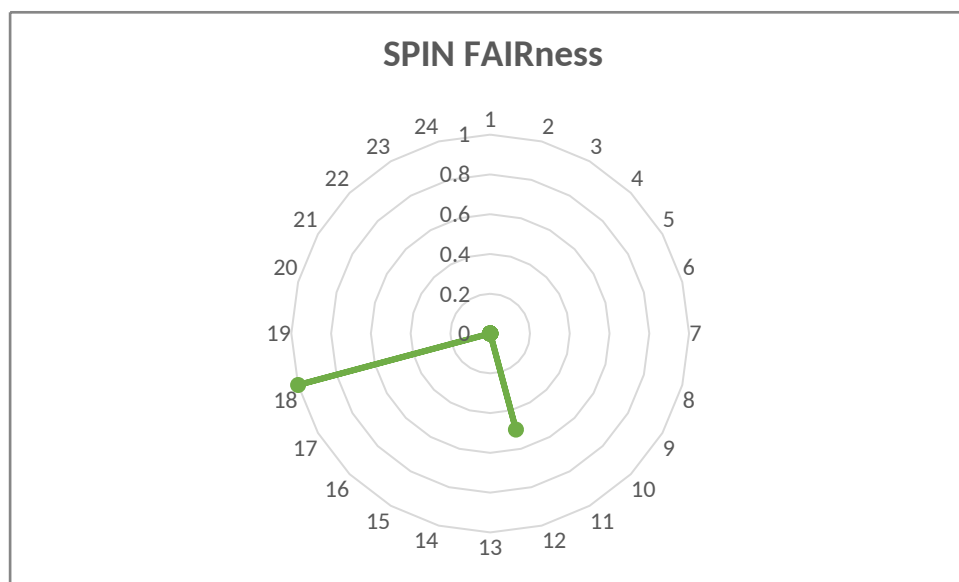


2.2.8. CNR-SPIN

Thin film fabrication and analysis

Principio	Situazione Globale	Score
-----------	--------------------	-------

F	Non viene usata una nomenclatura standard interna al laboratorio, e dati e metadati non sono pubblicati con un PID. I metadati vengono raccolti in un quaderno di laboratorio cartaceo fatta eccezione per quelli generati automaticamente dalle pipeline di acquisizione. Ai metadati non è associata una chiave univoca e la corrispondenza fra dati e metadati pubblicati è lasciata all'arbitrio del singolo operatore.	0,00
A	Dati e metadati non sono pubblicati e non viene mai dato accesso ai dati ad utenti esterni.	0,00
I	Fatta eccezione per i metadati salvati automaticamente dalla pipeline, i metadati sono in formato cartaceo e non pubblicati, quindi non interoperabili.	0,08
R	I metadati raccolti a mano su carta sono dettagliati, ma ogni riferimento ai dati non è standard e la ricostruzione dell'origine dei dati dipende dalle scelte del singolo operatore. In più non essendo pubblicati né i dati né i metadati non si fa uso di licenze e la riproducibilità e il riutilizzo dei dati è impossibile.	0,14
Media	Carenza : sia nella fase di raccolta che pubblicazione dei dati e metadati	0,06

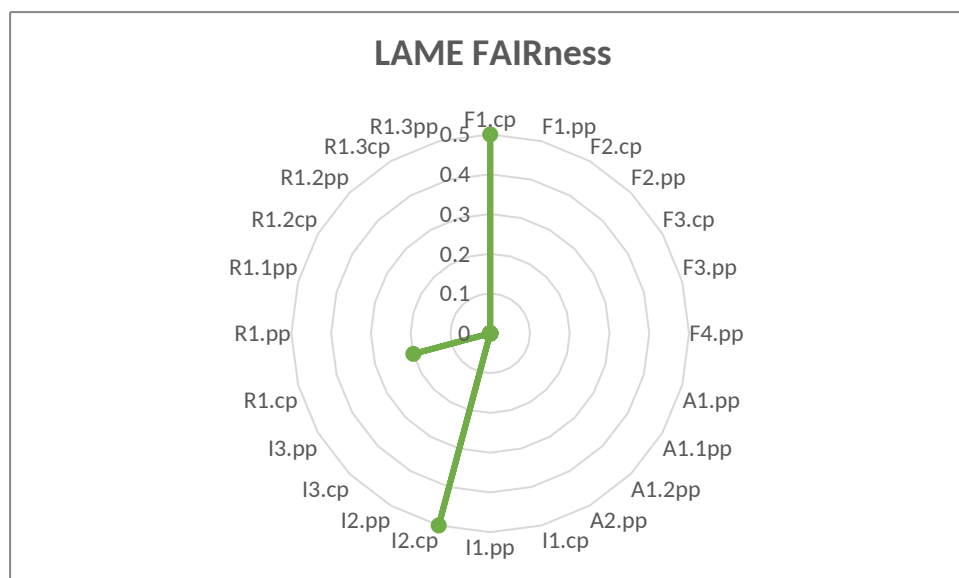


2.2.9. AREA SCIENCE PARK

LAME

Principio	Situazione Globale	Score
-----------	--------------------	-------

F	Dati raccolti con una nomenclatura standard ma non i metadati . Non si fa uso di un quaderno di laboratorio elettronico.	0,07
A	Dati salvati solo su server locali o cloud. Su richiesta può essere fornito accesso a utenti esterni.	0,00
I	I metadati non vengono raccolti da una pipeline automatica. Non vengono utilizzate ontologie specifiche ma è in corso lo studio di schemi di metadati standard. A livello locale il collegamento fra dati e sample è univoco.	0,08
R	Tutte le best practice relative alla riusabilità e riutilizzabilità dei dati e metadati sono in fase di studio e di progettazione ma al momento i metadati non sono raccolti, non si tiene traccia dei workflow né dell'origine dei dati e non essendo pubblicati i dati e i metadati non hanno licenze associate.	0,03
Media	Carenza: sia nella fase di raccolta che di pubblicazione di dati e metadati.	0,05

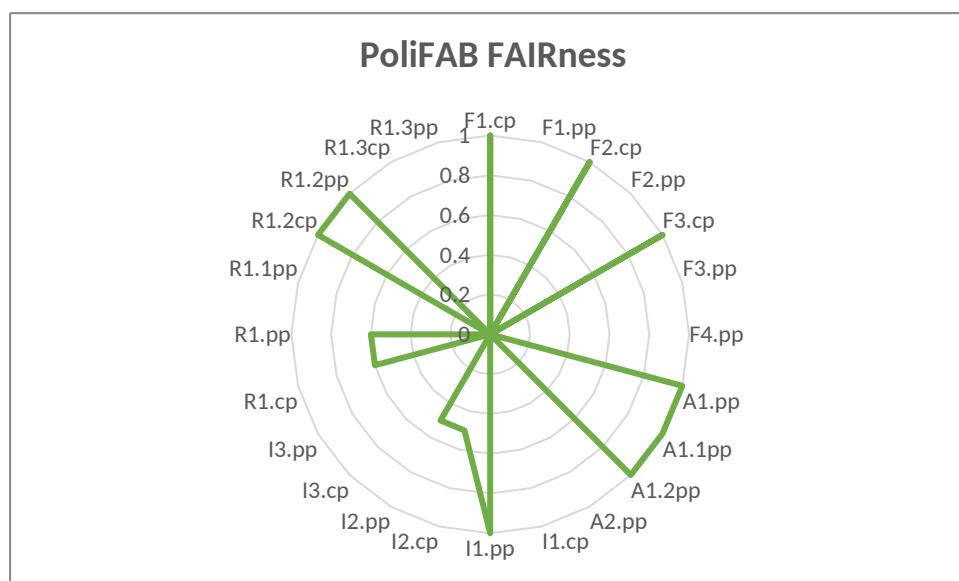


2.2.10. POLIFAB-POLIMI

PoliFAB

Principio	Situazione Globale	Score
-----------	--------------------	-------

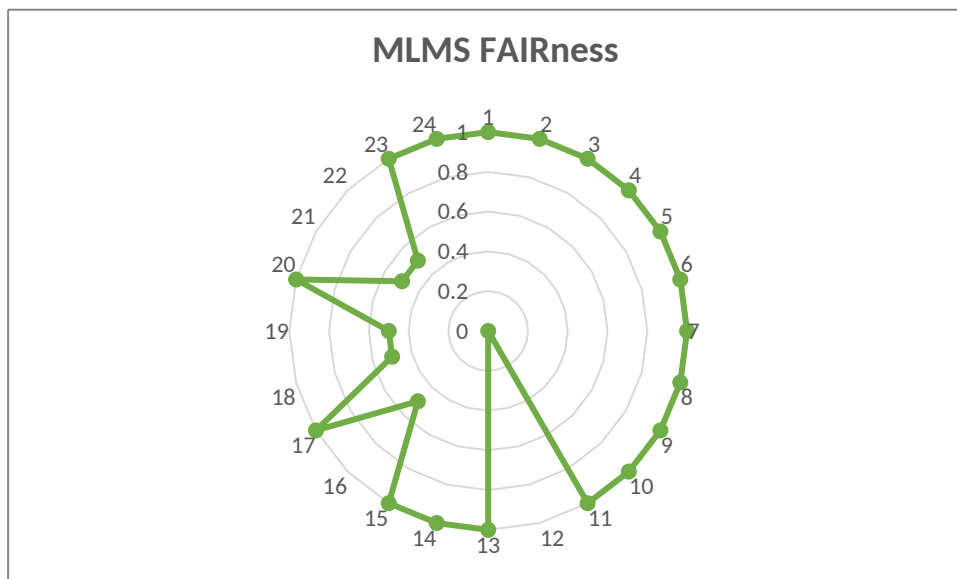
F	Dati e metadati raccolti direttamente dalla strumentazione, per cui è implementata una nomenclatura standard ma al momento i metadati non sono pubblicati con un PID. Si fa uso di eLabFTW come quaderno di laboratorio elettronico e in questo modo dati e metadati sono associati tramite identificativo univoco.	0,43
A	L'accesso ai dati e ai i metadati è gestito tramite eLabFTW, software open source e la cui istanza è resa pubblica.	0,75
I	I metadati non vengono raccolti da una pipeline automatica, ma sono disponibili in un formato standard. Non vengono utilizzate ontologie specifiche ma vengono però usate dei template definiti e documentati in eLabFTW. e non sono stati definiti schemi di metadati appositi.	0,33
R	I metadati sono raccolti in maniera esaustiva quasi completa, incluse le informazioni sul workflow, e viene ricostruita in parte l'origine dei dati. Tuttavia non vengono strutturati utilizzando schemi standard e vengono pubblicati facendo uso di licenze.	0,46
Media	Maggiore carenza: fase pubblicazione dei metadati.	0,47



2.2.11. UNIMI

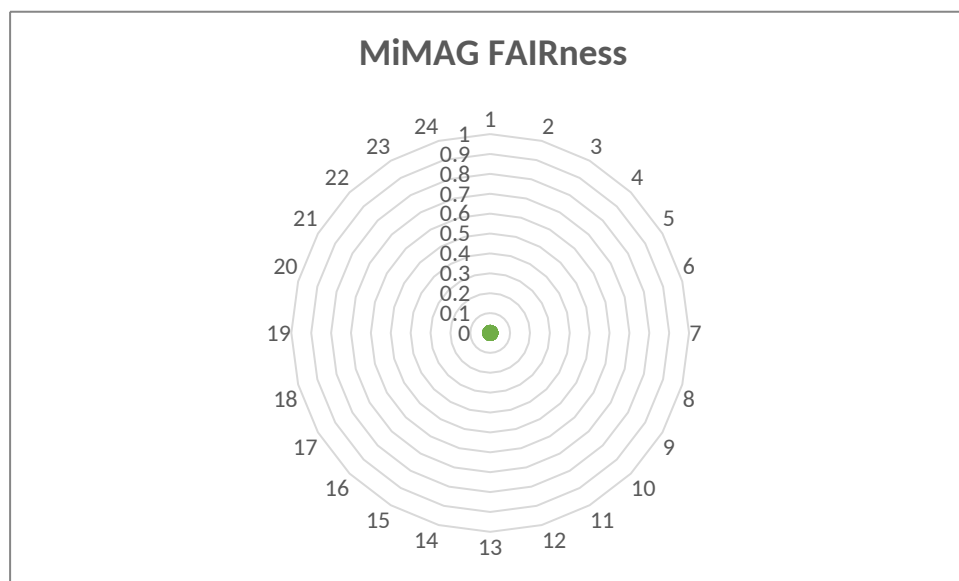
MLMS (Theory)

Principio	Situazione Globale	Score
F	Si tratta di un laboratorio teorico di simulazioni. Grazie a progetti precedenti hanno introdotto l'utilizzo del Data Management Plan. Fanno uso di nomenclatura standard, e nei file di input e output sono contenuti metadati rilevanti. Quando i dati sono pubblicati su AIR o il dataverse universitario vengono pubblicati anche i metadati a cui è associato un PID.	1,00
A	Dati e metadati pubblicati su AIR o sul dataverse universitario dove l'accesso è garantito dai protocolli di queste repository. I metadati vengono preservati.	1,00
I	I metadati sono salvati insieme ai dati e non vengono estratti da pipeline automatiche. Quando pubblicati usano un formato standard e seguono delle ontologie predefinite. Quasi sempre viene fatto chiaro riferimento ai dataset corrispondenti e ad altri metadati rilevanti.	0,75
R	Non sempre i metadati contengono tutte le informazioni relative ai workflow, per cui non sempre è ricostruibile la chiara origine dei dati. Dati e metadati sono pubblicati con licenze appropriate e seguendo ontologie riconosciute dalla comunità di riferimento.	0,71
Media	Unica carenza: automatizzare e ampliare raccolta dei metadati.	0,85



MiMAG (Experimental)

Principio	Situazione Globale	Score
F	La strumentazione è quasi tutta nuova e non ancora arrivata, ma al momento non è definita alcuna nomenclatura standard nè si sono definiti quali metadati salvare, dove pubblicarli o quale repository utilizzare	0,00
A	Dati e metadati verranno pubblicati e resi accessibili solo se ritenuti buoni. Non si è scelto però nessun repository o altro sistema di condivisione se non una cartella condivisa.	0,00
I	Trattandosi di strumentazione nuova non sono stati definiti format, ontologie, riferimenti ai dati corrispondenti all'interno dei metadati.	0,00
R	Trattandosi di strumentazione nuova non si è ancora deciso come inserire nei metadati tutte le informazioni necessarie al riutilizzo e alla riproducibilità dei dati.	0,00
Media	Carenza : sia nella fase di raccolta che pubblicazione dei dati e metadati	0,00



3. PIANO D'AZIONE

3.1. CLASSIFICAZIONE DELLE AZIONI NECESSARIE

L'analisi puntuale delle risposte ottenute ha evidenziato che i laboratori del progetto NFFA-DI, si possono raggruppare in insiemi distinti che condividono uno stato di FAIRness analogo e richiedono implementazioni simili. Qualitativamente abbiamo laboratori quasi FAIR, abbastanza FAIR e per niente FAIR. Per quantificare questa aggregazione partendo dai punteggi medi ottenuti da ciascun laboratorio, con un algoritmo di clustering gerarchico si è ottenuto un dendrogramma che si è tagliato applicando il metodo del gomito alla distanza di aggregazione e si è così individuato il valore di 0.2. In questo modo si sono ottenuti 4 gruppi corrispondenti alle soglie seguenti

- **Gruppo A**
 - punteggio medio >80%;
- **Gruppo B**
 - punteggio medio >30% e <=80% ;
- **Gruppo C**
 - punteggio medio >=9% e <30%;
- **Gruppo D**
 - punteggio medio <9% .

In base ai criteri sopra riportati i laboratori elencati nella Tabella 2 si distribuiscono come mostrato nella tabella che segue.

Nodo	Laboratorio	Tecniche	Ancillary Equipment	Gruppo FAIRness
CNR-IOM	APE HE beamline	XAS, XMCD	LEED	B
	APE Cluster	PES, MOKE	MBE, RHEED	/
	APE LE	PES	LEED	C
	INSPECT	PES	LEED	D

	STRAS - LEEM-PEEM	LEEM-PEEM	LEEM, PEEM	/
	STRAS - STM-VT	STM	PES, LEED	A
	STRAS - STM-LT	STM	LEED	A
	SPRINT - RAMAN	RAMAN		C
	SPRINT-TRANSIENT GRATING	TRANSIENT GRATING		C
	SPRINT - PHOTOLUMINESCENCE	PHOTOLUMINESCENCE		C
	SPRINT - Mott spectroscopy	TR-ARPES	PES, ARPES, SPIN-POL	C
	HMMBE	HMMBE	MBE, RHEED	B
	MBE/XPS	MBE/XPS		C
CNR-IFN@MI	UDYNI			C
CNR-IFN@TN	CleanRoom CRD @ FBK	LITHOGRAPHY & PATTERNING - RIE		B
CNR-IMM@BO				D
CNR-IMM@CT	Experimental	ATOM PROBE TOMOGRAPHY, STRUCTURAL AND ANALYTICAL IMAGING		C
	Theory - Simulation			B
CNR-ISM	EuroFEL Support Lab			B
CNR-NANOTEC	Clean Room E (nanofab)	SEM	FIB	C
CNR-SPIN				D
AREA SCIENCE PARK	LAME	TEM, SEM	FIB-SEM	D
POLIMI	PolifAB			B
UniMI	Exp MiMAG	SEM-PA, XPS, RAMAN		D
	Theory MLMS			A

Tabella 4: distribuzione dei laboratori rispetto ai 4 gruppi identificati.

Confrontando i giudizi qualitativi corrispondenti a laboratori appartenenti allo stesso gruppo assegnati dall'algoritmo si nota che l'aggregazione è coerente. Qui di seguito vengono riportate le principali azioni necessarie per ciascun gruppo.

3.1.1. GRUPPO A

Questi laboratori hanno già iniziato l'implementazione delle procedure FAIR-by-design e le azioni per raggiungere l'obiettivo del FAIR-by-design completo sono minime:

- concludere ed eventualmente estendere la definizione delle ontologie per i metadati;
- implementare la loro acquisizione automatica in maniera coerente.

3.1.2. GRUPPO B

Nei laboratori del Gruppo B sono già ben implementate svariate best practices necessarie per l'implementazione del FAIR-by-design. Si rileva tendenzialmente una mancanza di conformità nella fase di pubblicazione dei metadati.

Azioni necessarie per il raggiungimento dell'obiettivo:

- completare la raccolta di tutti i metadati (tutti tranne il CRD);
- definire delle ontologie per i metadati;
- implementare la loro acquisizione automatica in maniera coerente (tutti tranne il CRD);
- pubblicare i dati (CRD, HMMBE) e tutti i metadati.

3.1.3. GRUPPO C

Questo gruppo è particolarmente eterogeneo: su un base comune di totale non conformità nel trattare i metadati, a seconda dei casi alcuni laboratori sono carenti nella definizione delle nomenclature, altri nella pubblicazione stessa dei dati, altri nei formati con cui vengono salvati in parte i metadati. Differiscono dal GRUPPO D perché a seconda dei casi qualcosa è già impostato.

In generale le azioni necessarie per il raggiungimento dell'obiettivo sono:

- Introdurre un quaderno di laboratorio elettronico;
- creare e normalizzare pipeline di acquisizione dati e metadati;
- introdurre una nomenclatura interna se assente o incompleta
- definire delle ontologie per i metadati;
- scegliere dei repositories se non ancora fatto;
- pubblicare dati e metadati.

3.1.4. GRUPPO D

Per i laboratori di questo gruppo l'implementazione del FAIR-by-design parte da 0.

È pertanto necessario:

- definire nomenclature interne univoche e standard;
- introdurre un quaderno di laboratorio elettronico;
- creare e normalizzare pipeline di acquisizione dati e metadati;
- definire le ontologie per la raccolta dei metadati;
- scegliere i repository;
- pubblicare dati e metadati.

3.2. COORDINAMENTO CENTRALIZZATO

Nel WP3, l'attività 3.1 "FAIR Data Division" ha l'obiettivo di contribuire alla configurazione dell'infrastruttura dati e dei relativi servizi dati dell'ecosistema digitale OFED, che saranno messi a disposizione di tutto il progetto. Questo sarà realizzato all'interno del datacenter ORFEO installato nell'Area Science Park attraverso l'espansione dello storage e dell'hardware al

momento disponibili. L'infrastruttura di gestione dati è descritta in maniera dettagliata nel Deliverable D1.2 "Data Management Plan- First Release".

Come già anticipato nella Sezione 1 il "FAIR Data Management Team", introdotto nel D1.2 e riportato nella Tabella 1, fornirà un supporto strutturato a tutti i laboratori del progetto. Da Gennaio 2024 il team contatterà i vari laboratori per impostare una strategia di azione personalizzata per le implementazioni necessarie per raggiungere gli obiettivi di FAIRness. Si prevedono delle call ordinarie a cadenza bimestrale (eventualmente trimestrale per il laboratori del Gruppo A) per valutare lo stato di avanzamento dei lavori e delle call straordinarie ogni qualvolta si renda necessario. Inoltre, saranno programmate delle visite di persona ai laboratori per approfondire la conoscenza dei set-up sperimentali e le necessità specifiche delle singole realtà.

Nel D1.2 si sottolinea che i laboratori del progetto NFFA-DI sono tenuti ad utilizzare un quaderno elettronico di laboratorio integrato nel sistema OFED, qualora non si faccia già uso di un quaderno elettronico integrabile. Al momento è disponibile un'istanza di test del software open source [elabFTW](#) [4] e nei prossimi mesi verrà fornito accesso e supporto per il suo utilizzo a tutti i laboratori.

3.3. CREAZIONE GRUPPI DI LAVORO

Dal supporto fornito per la compilazione del form è emersa l'esigenza di creare dei gruppi di lavoro omogenei per tipologia di tecnica ed eterogenei quanto a FAIRness in modo da condividere scelte tecniche, ontologie e standard e in modo che i laboratori con un punteggio inferiore possano beneficiare da quanto già fatto e messo in pratica da altri laboratori analoghi.

Si è preparato un secondo form volto pertanto a integrare le informazioni raccolte col primo il cui scopo è di andare a completare la Tabella 4 e creare dei gruppi di lavoro sulla piattaforma di condivisione.

Il form è raggiungibile sempre attraverso la pagina web di [NFFA-DI](#) e al momento si stanno raccogliendo le risposte da parte dei Responsabili di Laboratorio. Per completezza anche questo questionario è riportato nell' Annex II.

4. OBIETTIVI FAIRNESS E CONSIDERAZIONI FINALI

Il quadro emerso dall'analisi delle risposte raccolte attraverso il questionario è molto eterogeneo. In alcuni casi i laboratori hanno già un'impostazione FAIR-by-design e sono richieste pochissime azioni per soddisfare i criteri richiesti, in altri casi è necessario un cambio profondo nel modo di lavorare e nella gestione dei dati e dei metadati.

Entro la fine del progetto gli obiettivi che il FAIR Data Management Team si pone sono di raggiungere il 95% di FAIRness per i Gruppi A e B, e di superare il 50% per i Gruppi C e D. Tali percentuali sono indubbiamente sfidanti soprattutto per quanto riguarda gli ultimi due gruppi.

Tuttavia, la creazione di gruppi di lavoro omogenei nelle tecniche sperimentali e la condivisione di tool, soluzioni, software permetteranno indubbiamente di velocizzare il processo di FAIRificazione dei laboratori appartenenti ai Gruppi C e D.

Per ultimo, ma non meno importante, il supporto nel continuo e la visita dei laboratori da parte del team riportato nella Tabella 1 permetterà di inquadrare meglio le criticità maggiori e di impostare una collaborazione proficua.

5. BIBLIOGRAFIA

[1] [NFFA-DI](#)

[2] <https://doi.org/10.1038/sdata.2016.18> (Wilkinson et al.)

[3] [FAIR principles translation SNSF logo \(snf.ch\)](#)

[4] [eLabFTW - free open source ELN](#)

6. ANNEXES

6.1. FAIRNESS STATUS MONITORING

- **Name**
- **Email Address**
- **Node**
- **Laboratory name**
- **F1.** (Meta)data are assigned a globally unique and persistent identifier - Data Collection Phase: during your experiment do you collect (meta)data by using a well defined convention that does not change along the experiment and does not change among experiments/projects?
- **F1.** (Meta)data are assigned a globally unique and persistent identifier - Data Published Phase: do you publish data and metadata with a PID?
- **5domande-F2.** Data are described with rich metadata (defined by R1 below) - Data Collection Phase: when you collect data, do you collect also all metadata associated in the different experimental phases? Do you use a laboratory notebook? Is it open? Do you collect sample's metadata? What about instruments metadata?
- **F2.** Data are described with rich metadata (defined by R1 below) - Data Published Phase: do you fully document each dataset in the metadata, which may include descriptive information about the context, quality and condition, or characteristics of the data. Another researcher in any field, or their computer, should be able to properly understand the nature of your dataset.
- **F3.** Metadata clearly and explicitly include the identifier of the data they describe - Data Collection Phase: on your server, do you use a unique key to link metadata with their corresponding data?
- **F3.** Metadata clearly and explicitly include the identifier of the data they describe - Data Published Phase: when you publish metadata, do they include the PID of the corresponding data?
- **F4.** (Meta)data are registered or indexed in a searchable resource - Data Published Phase: do you provide detailed and complete metadata for each dataset ?
- **A1.** (Meta)data are retrievable by their identifier using a standardised communications protocol - Data Published Phase: do you clearly define who can access the actual data? Do you specify how? It is possible that data will actually not be downloaded, but rather reused in situ. If so, the metadata must specify the conditions under which this is allowed (sometimes versus the conditions needed to fulfill for external usage/"download"). 3 domande

- **A1.1** The protocol is open, free, and universally implementable - Data Published Phase: may anyone with a computer and an internet connection access at least the metadata ?
- **A1.2** The protocol allows for an authentication and authorisation procedure, where necessary - Data Published Phase: how may external users have access to your data? It often makes sense to request users to create a user account on a repository. This allows to authenticate the owner (or contributor) of each dataset, and to potentially set user specific rights.
- **A2.** Metadata are accessible, even when the data are no longer available - Data Published Phase: do you preserve metadata ?
- **I1.** (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. - Data Collection Phase: when you collect metadata, do you use automatic pipeline to save them in a standard format that can be understood both by humans and machines?
- **I1.** (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. - Data Published Phase: do you publish metadata in a standard format that can be understood both by humans and machines?
- **I2.** (Meta)data use vocabularies that follow FAIR principles - Data Collection Phase: are the vocabularies/ontologies/thesauri you use findable, accessible, interoperable and thoroughly documented, hence FAIR?
- **I2.** (Meta)data use vocabularies that follow FAIR principles - Data Published Phase: are the vocabularies/ontologies/thesauri you use findable, accessible, interoperable and thoroughly documented, hence FAIR? Researchers can refer to metrics assessing the FAIRness of a digital resource (if available).
- **I3.** (Meta)data include qualified references to other (meta)data - Data Collection Phase: do you properly cite relevant/associated datasets, in particular by providing their name convention, in the metadata, and describe the scientific link/relation to your dataset ?
- **I3.** (Meta)data include qualified references to other (meta)data - Data Published Phase: when you publish your data&metadata, do you properly cite relevant/associated datasets, in particular by providing their persistent identifiers, in the metadata, and describe the scientific link/relation to your dataset?
- **R1.** (Meta)data are richly described with a plurality of accurate and relevant attributes - Data Collection Phase: do you provide complete metadata for each data file? Some points to take into consideration (non-exhaustive list): a) Scope of your data: for what purpose was it generated/collected? b) Particularities or limitations about the data that other users should be aware of. c) Date of the dataset generation, lab conditions, who prepared the data, parameter settings, name and version of the software used. d) Is it raw or processed data? e) Variable names are explained or self-explanatory (i.e. defined in the research field's controlled vocabulary).
- **R1.** (Meta)data are richly described with a plurality of accurate and relevant attributes - Data Published Phase: do you provide complete metadata for each data file? Some points to take into consideration (non-exhaustive list): a) Scope of your data: for what purpose was it generated/collected? b) Particularities or limitations about the data that other users should be aware of. c) Date of the dataset generation, lab conditions, who prepared the data, parameter settings, name and version of the software used. d) Is it raw or processed data? e) Variable names are explained or self-explanatory (i.e. defined in the research field's controlled vocabulary) f) Version of the archived and/or reused data is clearly specified and documented.

- **R1.1.** (Meta)data are released with a clear and accessible data usage license - Data Published Phase: do you include information about the license in the metadata? If a particular license is needed, you have to provide it along with the dataset. Where possible it is suggested to use creative common licenses, such as CC0, CC BY, etc., which can be referred to by URL.
- **R1.2.** (Meta)data are associated with detailed provenance - Data Collection Phase: along the data collection phase, do you save all the metadata that describe the workflow that led to your data?
- **R1.2.** (Meta)data are associated with detailed provenance - Data Published Phase: do your metadata describe the workflow that led to your data? Who generated or collected it? How has it been processed? Has it been published before? Does it contain data from someone else, potentially transformed or completed? Ideally the workflow is described in a machine-readable format. Criterion I3 is closely linked to this issue when reusing published datasets.
- **R1.3.** (Meta)data meet domain-relevant community standards - Data Collection Phase: do you prepare your (meta)data according to community standards and best practices for data archiving and sharing in your research field?
- **R1.3.** (Meta)data meet domain-relevant community standards - Data Published Phase: do you publish metadata according to community standards and best practices for data archiving and sharing in your research field? There might be situations where good practice exist for the type of data to be submitted but the submitter has valid and specified reasons to divert from the standard practice. This needs to be addressed in the metadata.

6.2. CENSIMENTO DI TECNICHE E STRUMENTI A CATALOGO NFFA-DI

- **Operative Unit**
- **Instrument name**
- **Facility type**
- Experimental technique for NFFA-DI catalogue
- New experimental technique
- Technique sub-category
- Ancillary equipment
- Person in charge of the technique on this Instrument
- Email Address
- Raw data file format
- You selected some proprietary file format, not FAIR compliant. To which open file format will you convert your data?
- Other file format - File extension (without dot)
- Other file format - Is it an open format (or at least editable with an open software)?
- Other file format - You inserted some proprietary file format, not FAIR compliant. To which file format will you convert your data?
- Average amount of data produced in 8 hours (MB)
- Does the instrument allow for remote access?
- If the instrument allows the remote access, which software do you use to do it?