

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.

D1.2 – DATA MANAGEMENT PLAN FIRST RELEASE

BIM 5

OI1.2 – NFFA-DI data management plan adoption

WP1 – Management

NANO FOUNDRIES AND FINE ANALYSIS DIGITAL INFRASTRUCTURE
NFFA-DI



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>
27/10/23	Draft	<i>Irene Modolo, Federica Bazzocchi, Ruggero Lot, Dario De Angelis, Mirco Panighel, Cecilia Zagni</i>		
31/10/23	Draft	<i>Irene Modolo, Federica Bazzocchi, Ruggero Lot, Dario De Angelis, Mirco Panighel, Cecilia Zagni</i>		Stefano Cozzini
14/12/23	v.1.0.0	<i>Federica Bazzocchi Mariarita de Luca</i>		Stefano Cozzini

Sommario

1. INTRODUZIONE; IL PROGETTO NFFA-DI.....	4
1.1. Livelli di gestione dei Dati all'interno di NFFA-DI.....	4
2. TIPOLOGIA DEI DATI PRODOTTI E GESTITI NEL PROGETTO.....	5
2.1. Descrizione dei dati gestiti.....	5
2.2. L'infrastruttura: OFED e gli strumenti utilizzati.....	8
2.3. Il flusso dei dati in OFED.....	10
2.4. DMP di Laboratorio.....	12
3. FAIR-by-design.....	12
3.1. Dati FAIR.....	12
3.1.1. Findability: rendere i dati reperibili includendo metadati appropriati.....	13
3.1.2. Accessibility: rendere i dati accessibili.....	14
3.1.3. Interoperability: rendere i dati interoperabili.....	14
3.1.4. Reusability: aumentare il riutilizzo dei dati.....	15
3.2. FAIR DATA MANGEMENT TEAM.....	15
4. Altri prodotti della ricerca.....	16
5. Allocazione delle risorse.....	16
6. Sicurezza dei dati.....	17
6.1. Policy di Backup.....	17
6.2. Conservazione dei dati.....	17
7. Aspetti etici.....	18
8. Annex 1: Politica in materia dei dati di ricerca NFFA-DI.....	18
9. Annex 2: Glossario.....	20

ABSTRACT

Il presente documento rappresenta il deliverable D1.2 “Data Management ” del progetto NFFA-DI, redatto nell’ambito del Work Package 1 (WP1) corrispondente al “Management” del progetto e definisce le modalità di gestione di tutti i Dati prodotti nell’ambito del progetto NFFA-DI. Questi si dividono in tre categorie:

- Documentazione e Relazioni di progetto;
- Dati di Ricerca;
- Dati Personali degli utenti registrati sul sito <https://nffa-di.it/> necessari alla gestione del progetto stesso.

Lo scopo di questo documento è definire una strategia comune relativa alla gestione dei diversi tipi di Dati durante l’intero ciclo di vita del progetto.

1. INTRODUZIONE; IL PROGETTO NFFA-DI

Il progetto NFFA-DI mira a realizzare un’infrastruttura di ricerca distribuita su scala nazionale per le nanoscienze e le nanotecnologie, con l’obiettivo di creare un ambiente unico sia per le nanoscienze di base che per le tecnologie avanzate, colmando il divario tra la ricerca fondamentale sulla materia quantistica e i sistemi micrometrici funzionali alla transizione digitale. La combinazione dell’accesso digitale ad un ampio portfolio di servizi tramite la realizzazione di un Portale Unico e di un catalogo di risorse sperimentali e computazionali avanzate, assieme ai servizi basati sui dati FAIR, sostiene l’intera filiera della ricerca, dalla scoperta di nuovi materiali al trasferimento tecnologico verso l’industria.

NFFA-DI si prefigge di diventare un’infrastruttura di ricerca di riferimento nel panorama europeo, possedendo un ottimo potenziale per l’accelerazione della transizione digitale nella ricerca e per la società.

Il progetto prevede l’acquisto e l’aggiornamento di strumentazione sperimentale di ultima generazione per il rafforzamento dei laboratori afferenti alle varie “Unità Operative” (UO) del progetto e l’implementazione e l’utilizzo di specifici software open source per la gestione dei dati in modalità FAIR-by-design. Questo progetto è molto ambizioso data l’eterogeneità dei laboratori e delle tecniche sperimentali coinvolte.

L’ampio spettro di strumenti scientifici e tecniche di misura disponibili nei siti partner sarà accessibile da parte di ricercatori accademici e industriali in seguito alla presentazione e all’approvazione di una proposta di ricerca (*Proposal*) attraverso un Portale Unico.

I dati scientifici prodotti nell’ambito del progetto NFFA-DI saranno dunque generati dall’attività di ricerca sia dei ricercatori affiliati alle UO sia da quella degli utenti esterni che avranno accesso all’infrastruttura.

1.1. Livelli di gestione dei Dati all’interno di NFFA-DI

Questo documento rappresenta il Data Management Plan (DMP) del progetto NFFA-DI nel suo complesso, con lo scopo di definire le linee guida nella gestione dei dati del progetto stesso che saranno successivamente recepite da ogni laboratorio.

Vi vengono inoltre definiti i ruoli e le responsabilità dei team di progetto dedicati al supporto e al coordinamento del FAIR-by-design.

Per una panoramica degli obblighi e delle responsabilità a cui sono soggetti gli utilizzatori della strumentazione NFFA-DI in tema di gestione dei Dati della Ricerca si rinvia all’ Annex 1 (*Politica in materia dei Dati di Ricerca NFFA-DI*).

Questo DMP è, come detto, un documento che coordina la gestione dei dati a livello macro dell'intero progetto in maniera globale. La gestione dei dati in fase di raccolta è gestita a livello dei DMP di laboratorio (Lab-DMP) che rifletteranno pertanto le specificità dei singoli laboratori.

A questo secondo livello saranno delineate le strategie e le modalità di gestione dei dati specifiche per la strumentazione relativa al laboratorio stesso, i ruoli e le responsabilità delle persone incaricate della corretta gestione dei dati degli esperimenti in modalità FAIR-by-design. Anche i Lab-DMP devono essere considerati *living document*, aggiornati ogni qualvolta sia necessario apportare delle modifiche o dei miglioramenti nella metodologia relativa alla gestione dei dati, alle tipologie di dati registrati, o ai software utilizzati.

Vi è infine un terzo livello di DMP, relativo ai progetti di ricerca che saranno sottomessi attraverso il Portale Unico di accesso all'infrastruttura di NFFA-DI e ad ogni progetto sarà associato un DMP. Qui di seguito indicheremo questi ultimi DMP come Proposal-DMP.

I Principal Investigator (PI) compileranno e sottometteranno un Proposal-DMP relativo ad ogni singolo progetto di ricerca sottomesso e approvato, facendo riferimento ai Lab-DMP dei laboratori coinvolti nel progetto. Auspicabilmente, per facilitare gli utenti, tale Proposal-DMP sarà redatto per mezzo di un form sul portale NFFA-DI che risulterà già completo delle informazioni relative al formato dei dati grezzi e ai metadati acquisiti in base alla strumentazione che andranno ad utilizzare, ma che richiederà loro dettagli aggiuntivi strettamente legati alla gestione dei dati inerenti allo specifico progetto di ricerca¹.

Ogni PI dovrà poter modificare e salvare il Proposal-DMP durante il corso del proprio progetto, e finalizzarlo entro un mese dalla data di conclusione degli accessi programmati presso l'infrastruttura².

Infine, il presente DMP va considerato un documento vivo (*living document*), che sarà aggiornato, quando necessario, durante l'intero ciclo di vita del progetto. Tutte le versioni successive saranno tracciate e accessibili tramite un Persistent Identifier (PID).

2. TIPOLOGIA DEI DATI PRODOTTI E GESTITI NEL PROGETTO

2.1. Descrizione dei dati gestiti

Come spiegato nell'Introduzione, questo DMP tratta la gestione dei dati raccolti all'interno del progetto NFFA-DI. Le tipologie di dati raccolti sono le seguenti:

- Documentazione e relazioni di progetto: in questo caso la maggior parte dei dati raccolti consisterà in vari tipi di documenti. Ciò include informazioni amministrative, requisiti tecnici, relazioni e altra documentazione pertinente relativa alle attività e ai risultati del progetto;
- Dati di Ricerca: comprensivi dei dati stessi e dei relativi metadati;
- Dati Personali: i dati personali saranno raccolti tramite moduli di registrazione come parte delle attività del progetto. Ciò potrebbe includere informazioni personali come nomi, dettagli di contatto, affiliazioni e altri dati pertinenti forniti dalle persone durante il processo di registrazione.

¹ punto 4.4.2 della *Politica in materia dei Dati di Ricerca NFFA-DI*

² punto 4.4.3 della *Politica in materia dei Dati di Ricerca NFFA-DI*

Per ciascuna delle categorie sopra identificate viene presentata qui di seguito una tabella dove sono descritte alcune loro caratteristiche.

Documentazione e relazioni di progetto	
Scopo della raccolta/generazione dei dati	Documenti ufficiali, deliverable, milestone, relazioni di incontri, statistiche e altri documenti. La documentazione del progetto è condivisa tra le diverse componenti dei vari assi e pacchetti di lavoro in uno spazio dedicato, al momento sulla piattaforma NFFA Datashare (https://datashare.nffa.eu) e probabilmente in futuro su una piattaforma analoga dedicata al progetto NFFA-DI. Le autorizzazioni saranno gestite in modo dettagliato su cartelle individuali e anche su documenti individuali, per garantire un accesso ai dati coerente con i ruoli assegnati all'interno del progetto NFFA-DI.
Formato e tipologia di dati	Preferibilmente formati di file aperti, anche se sono ammessi alcuni file in formati proprietari quali .docx, .xlsx, purché modificabili tramite programmi open source.
Origine dei dati	La documentazione del progetto, prodotta da personale tecnico/amministrativo e da persone coinvolte nel progetto.
Volume atteso	Meno di 500 GB.
Utilità dei dati	Questi dati vengono utilizzati per monitorare l'andamento del progetto, seguire il processo di approvazione delle proposte e per ottemperare a qualsiasi obbligo legale, contrattuale e normativo in relazione al PNRR, ai revisori e agli organi statutari e di verifica aziendali.
Work Package di riferimento	WP1
Gestione	CNR

Tabella 1: schema descrittivo dei dati di tipo documentale.

Dati di Ricerca	
Scopo della raccolta/generazione dei dati	I Dati di Ricerca vengono generati, raccolti, processati e analizzati per attività di ricerca specifiche, che nella maggioranza dei casi porteranno a delle Pubblicazioni Scientifiche.
Formato e tipologia di dati	<p>Data la grande varietà di strumenti scientifici e tecniche di misura disponibili, nel corso del progetto verranno utilizzati vari formati di dati. In ogni caso, i dati necessari a validare i risultati presentati in una Pubblicazione Scientifica saranno convertiti in un formato file che possa essere aperto con un software multiplatforma open source (o almeno gratuito) in modo da consentire a terzi di accedervi, estrarli, sfruttarli, riprodurli e diffonderli gratuitamente, rispettando ovviamente le licenze con cui verranno pubblicati.</p> <p>Nella versione successiva di questo documento verranno fornite informazioni più dettagliate su formati e tipi di dati generati. A questo scopo, i ricercatori a capo dei laboratori (Responsabili di Laboratorio) che aderiscono al progetto NFFA-DI sono tenuti a fornire, per ogni strumento scientifico facente capo al progetto, le informazioni sul formato di dati grezzi e metadati acquisiti, nonché le informazioni su eventuali software di analisi dati utilizzati e formati esportabili dei dati analizzati sottomettendo un DMP di laboratorio (Lab-DMP) https://nffa-di.it/labequipment.</p>
Origine dei dati	I dati sono generati dalla strumentazione acquistata, implementata o integrata nell'ambito del progetto NFFA-DI presso le varie unità operative.
Volume atteso	Basandoci sul numero di unità operative coinvolte e sulla tipologia della strumentazione acquistata o implementata, per i prossimi 12 mesi è attesa una mole di dati compresa tra i 10 e i 50 TB. Un volume più preciso sarà ricalcolato nelle versioni seguenti di questo documento man mano che le varie strumentazioni acquistate all'interno del progetto verranno installate e testate.
Utilità dei dati	I dati prodotti hanno rilevanza nell'ambito della ricerca scientifica, favorendo l'accelerazione della stessa e la formazione di personale qualificato. Allo stesso tempo sono utili a convalidare, revisionare e riprodurre i risultati scientifici già pubblicati in articoli di settore.
Work Package di riferimento	WP3
Gestione	Area Science Park

Tabella 2: schema descrittivo dei dati di ricerca.

Dati Personali	
Scopo della raccolta/generazione dei dati	Ragioni di gestione relative al progetto e alla corretta raccolta dei metadati di ricerca e gestione dell'accesso al sito web www.nffa-di.it e al Portale Unico per la sottomissione dei Proposal. Il trattamento dei dati personali necessari a registrarsi sul sito web www.nffa-di.it è disciplinato dalla privacy policy accessibile al link https://nffa-di.it/privacy-policy .
Formato e tipologia di dati	Formato testuale
Origine dei dati	Contenuti nella documentazione relativa al proposal, nei metadati dei dati di ricerca e nel database del sito web.
Volume atteso	Meno di 50 GB.
Utilità dei dati	I dati hanno utilità interna al progetto o legata alla corretta attribuzione dei risultati scientifici prodotti.
Work Package di riferimento	WP2
Gestione	UniMI

Tabella 3: schema descrittivo dei dati personali.

2.2. L' infrastruttura: OFED e gli strumenti utilizzati

Il ruolo dell'infrastruttura che verrà a crearsi per la gestione dei dati di ricerca è centrale all'interno del progetto NFFA-DI.

Il progetto ha previsto l'investimento di 1,7 M€ di hardware per effettuare l'upgrade di ORFEO, il Data Center di Area Science Park situato presso il Campus di Basovizza.

Gli acquisti che verranno effettuati con i fondi ottenuti nel progetto NFFA-DI intendono ampliare e completare l'infrastruttura di calcolo. Si prevede in particolare di acquistare alcune componenti hardware e una componente di servizi come dettagliato nel seguito:

1. Espansione Partizione HPC (PH): si intende espandere l'attuale capacità di calcolo aggiungendo nodi calcolo HPC senza acceleratori;
2. Espansione ed Integrazione Rete DataCenter (RD): si intende adeguare la capacità della rete di interconnessione interna al datacenter per l'inserimento dei nuovi nodi di calcolo mantenendo un margine di espansione, simile a quello attualmente per la connessione di apparecchiature future;
3. Data center in container (DCC): si intende acquistare una nuova soluzione di datacenter in container che ospiterà sia tutta la parte hardware dell'infrastruttura esistente che la parte hardware dell'infrastruttura oggetto di questo capitolato (infrastruttura fornita);

4. Servizi di installazione hardware, di migrazione e di integrazione hardware e software tra l'infrastruttura esistente e l'infrastruttura fornita nel nuovo datacenter. Gli incrementi hardware nelle quattro componenti sopra identificate richiedono la piena integrazione e compatibilità con l'infrastruttura esistente sia dal punto di vista hardware che da quello del software.

L'archiviazione di grandi quantità di dati scientifici in maniera sicura e affidabile pone molte sfide ai sistemi di archiviazione monolitici tradizionali, le più rilevanti – nel contesto del progetto NFFA-DI – sono scalabilità, disponibilità e bilanciamento del carico. Scalabilità perché, durante il corso del progetto e oltre, sarà necessario gestire quantità sempre crescenti di dati senza compromettere le prestazioni o l'affidabilità, problema che non può essere affrontato in maniera tradizionale con sola scalabilità verticale. Disponibilità e bilanciamento del carico perché un unico accesso ai dati si comporta come collo di bottiglia e singolo punto di rottura per l'intera infrastruttura.

Per queste ragioni, sarà adottata una soluzione basata su un'infrastruttura distribuita, ovvero un sistema costituito da più nodi (o server) che comunicano e cooperano tra loro per archiviare i dati. L'infrastruttura dedicata allo storage grazie agli acquisti pianificati sarà composta da almeno 1 PB di spazio RAW su dischi rotativi e 50 TB su supporti allo stato solido. Tale infrastruttura verrà distribuita su almeno 10 macchine e verrà distribuita con un meccanismo a bucket.

Gli investimenti in hardware e la ricerca e lo sviluppo di software appositi porterà alla realizzazione di un ecosistema digitale per la gestione dei dati in modalità FAIR-by-design, denominato OFED acronimo di *Overarching Fair Ecosystem for Data system*.

Per essere conformi ai principi FAIR, sfruttando le competenze e gli strumenti presenti nella comunità internazionale della scienza dei materiali, verranno proposti alcuni tool messi a disposizione dal consorzio tedesco FAIRmat. In particolare, l'Area B – *Experiment* (<https://www.fairmat-nfdi.eu/fairmat/areas-fairmat/area-b-fairmat>) del consorzio FAIRmat, tra le altre cose, si occupa di:

- descrivere in dettaglio una serie di metodi sperimentali consolidati, definendone una caratterizzazione definitiva, in modo tale che i dati di ricerca contengano sistematicamente tutti i dati e i metadati necessari per una comparabilità dettagliata degli esperimenti;
- sviluppare schemi di dati e metadati specifici per il settore, promuovendo la standardizzazione degli stessi e il collegamento con i quaderni di laboratorio elettronici.

In questa visione, i dati di ricerca verranno pubblicati su NOMAD (<https://nomad-lab.eu/nomad-lab>), una piattaforma di gestione dei dati (Data Repository), sviluppata da FAIRmat, gratuita e open source per la scienza dei materiali, che permette di pubblicare i dati e i metadati della ricerca scientifica in maniera strutturata e aderente ai principi FAIR.

NOMAD Oasis è un tool che consente di creare una versione locale di NOMAD, basata sullo stesso software di NOMAD. Poiché in NOMAD Oasis sono replicate tutte le principali funzionalità di NOMAD relative alla gestione dei dati, esso consente di implementare nuove modalità basate su esigenze locali. Ciò permette allo stesso tempo di gestire i dati localmente e in modo strutturato secondo i principi FAIR, ma anche di predisporre i dati stessi per la pubblicazione sulla piattaforma cloud NOMAD.

Alla piattaforma NOMAD Oasis gestita da NFFA-DI verrà collegato il quaderno di laboratorio elettronico eLabFTW, anch'esso gratuito e open source. Tale strumento faciliterà la raccolta dei metadati e la gestione degli esperimenti in laboratorio.

Il Work Package 3 (WP3), *Digital Structure of NFFA-DI and Overarching Fair Ecosystem for Data system (OFED)*, è interamente dedicato a questo compito impegnativo e mira a consolidare ed estendere quanto raggiunto nell'ambito dell'infrastruttura di ricerca NFFA-Europe (www.nffa.eu)

e a fornire agli utenti, ai ricercatori affiliati e ai partner gli strumenti necessari a gestire i dati da loro prodotti con una metodologia FAIR-by-design³.

2.3. Il flusso dei dati in OFED

OFED è pensato come un **Data Lake** per lo storage di dati e metadati in modalità FAIR-by-design e per fornire degli strumenti avanzati di analisi dati.

Nell'immagine seguente è rappresentata una panoramica completa del processo di data-flow verso OFED.

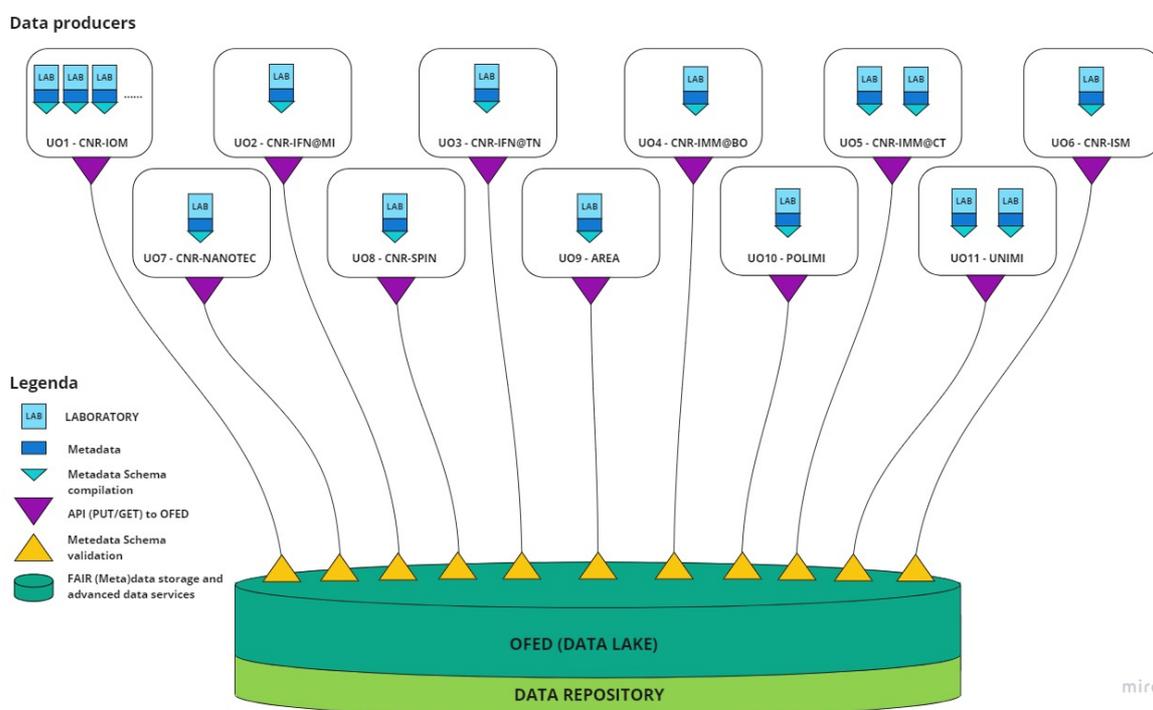


Figura 1: Schema del processo di acquisizione dei dati prodotti da ogni Laboratorio afferente alle varie Unità Operative di NFFA-DI verso OFED.

Come si evince dalla Figura 1, il flusso dei dati prevede tre livelli:

- Livello delle UO: i dati grezzi (raw data) sono generati dagli strumenti scientifici del laboratorio e raccolti sui database dei server locali. Gli operatori delle singole UO raccolgono i dati e compilano lo schema dei metadati sul notebook elettronico (elabFTW o altro software interfacciato a OFED), come dichiarato nel Lab-DMP. Successivamente, i dati vengono convalidati e inviati al Data Lake. In questo processo vengono attivati protocolli di sicurezza dei dati, come controlli di autenticazione e privacy.
- Livello del Data Lake: dati e metadati prodotti dalle UO si incontrano nel Data Lake dopo essere stati verificati e, se necessario, armonizzati secondo una politica conforme agli standard FAIR. Qui avviene anche la rielaborazione dei dati, i cui risultati vengono salvati su NOMAD Oasis.

³ NFFA-DI – Testo del progetto post-negoziazione

- Livello del Data Repository (piattaforma): dati finali e metadati vengono pubblicati su un Data Repository certificato in grado di renderli non più modificabili e di assegnare loro dei Persistent Identifiers (consigliabile NOMAD). Pubblicazioni scientifiche e documenti associati o altri documenti, ai quali è analogamente assegnato un PID, saranno invece pubblicati su Zenodo o altra repository analoga.

Per completezza e 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. 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 è definito *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 meno, ma che non rappresentano la motivazione per cui un utente sceglie di accedere all'infrastruttura.

UO	Laboratorio	Tecniche	Strumento	Attrezzatura Ancillare	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
	STRAS - STM-LT	STM		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	MBE		MBE	Silvia Rubini
In-silico experimental platform				Layla Martin Samos	
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	Experimental	TEM, SEM		FIB-SEM	Federica Bazzocchi
Polifab-POLIMI					Matteo Cantoni
UniMI	PolifAB				Andrea Giugni
	MLMS				Giovanni Onida

Tabella 4: dettaglio dei laboratori del progetto NFFA-DI.

2.4. DMP di Laboratorio

Prima della raccolta di qualsiasi Dato di Ricerca, i Responsabili di Laboratorio elencati nella tabella 4 sono tenuti a redigere un DMP (Lab-DMP) relativo alla gestione dei Dati di Ricerca prodotti durante il progetto. Il Lab-DMP deve essere redatto, e aggiornato quando necessario, utilizzando un modello comune.

I Lab-DMP prodotti da ciascun laboratorio verranno citati nelle versioni successive di questo documento. In caso di discrepanza o disaccordo tra un Lab-DMP e il presente documento (NFFA-DI DMP), NFFA-DI DMP avrà la precedenza.

3. FAIR-BY-DESIGN

3.1. Dati FAIR

NFFA-DI farà in modo di rendere i dati di ricerca prodotti dalla strumentazione coinvolta nel progetto reperibili, accessibili, interoperabili e riutilizzabili (FAIR), oltre a garantirne l'accesso aperto qualora questi fossero rilevanti per i risultati riportati in una pubblicazione scientifica. Questo è richiesto per essere in linea con gli obiettivi del Piano Nazionale per la Scienza Aperta, quelli di CoARA e dell'European Open Science Cloud⁴.

Lo schema di implementazione pratica dei principi FAIR è riportato nella tabella che segue:

Findable	Identificativo univoco e persistente
Accessible	(Meta)dati in una repository attendibile Accesso ristretto se necessario
Interoperable	Utilizzo di vocabolari e ontologie controllati Formati aperti
Reusable	Ricca documentazione Licenze di utilizzo chiare

Tabella 5: schema riassuntivo dell'implementazione dei principi FAIR.

⁴ punto 4.2.1 e 4.3.1 della *Politica in materia dei Dati di Ricerca NFFA-DI*

3.1.1. Findability: rendere i dati reperibili includendo metadati appropriati

Tutti i dati di ricerca e metadati associati generati dalla strumentazione acquistata, implementata o integrata nel contesto del progetto NFFA-DI saranno identificabili e localizzabili mediante un identificatore univoco universale (UUID) assegnato loro dal sistema OFED⁵.

Una parte di questi dati, ovvero quelli presenti o discussi in pubblicazioni scientifiche e presentazioni a conferenze, saranno resi disponibili ad accesso aperto entro la data di diffusione della corrispettiva pubblicazione o poster, utilizzando una data repository aperta in grado di assegnare loro un Persistent Identifier (PID)⁶. Al momento, la data repository di elezione è NOMAD (<https://nomad-lab.eu/nomad-lab>).

Obbligatoriamente, i metadati dei dataset pubblicati includeranno i seguenti metadati amministrativi⁷:

- Breve descrizione, Data di inserimento nella repository e Autore(i) del Dataset;
- Finanziamento: NextGenerationEU - PNRR "NFFA-DI" IR0000015 - 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";
- Nome, acronimo e numero del progetto di sovvenzione: "NANO FOUNDRIES AND FINE ANALYSIS - DIGITAL INFRASTRUCTURE, NFFA-DI, IR0000015";
- Termini di licenza: CC BY o una licenza con diritti equivalenti;
- Numero ID del proposal (se applicabile);
- PID delle pubblicazioni correlate e di altri prodotti di ricerca (se applicabile).

Il progetto NFFA-DI comprende svariate tecniche di misura in un ampio spettro di discipline scientifiche. È chiaramente impossibile identificare un unico schema di metadati (*Metadata Schema*) che sia esauriente per descrivere i parametri di ciascun esperimento.

Pertanto, è necessario redigere una serie di *Metadata Schema* che siano allo stesso tempo generali, per inglobare più tecniche di misura insieme, ma specifici, in modo da poter soddisfare le necessità delle singole tecniche di misura. Questo può essere fatto solo grazie a una stretta collaborazione fra i leader dei nodi e il team di data management.

La definizione e l'acquisizione di metadati associati a strumenti scientifici e tecniche di misura richiedono un forte impegno e collaborazione da parte dei gruppi di ricerca. Pertanto, tutti gli utilizzatori della strumentazione acquistata, implementata o integrata nel contesto del progetto NFFA-DI sono tenuti ad utilizzare un quaderno elettronico di laboratorio integrato nel sistema OFED, al fine di facilitare le buone pratiche di gestione dati, la condivisione di dati e documenti tra i ricercatori, risalire alla provenienza dei dati e proteggersi dalla loro perdita⁸.

Uno degli obiettivi principali del WP3 è l'implementazione di un approccio FAIR-by-design in tutte le Unità Operative del progetto NFFA-DI. Per far questo verrà fornito supporto ai vari laboratori per identificare e definire le procedure che portino al pieno controllo della gestione della provenienza dei dati e dei metadati associati.

In particolare, verranno elaborate e implementate procedure e raccomandazioni orientate al FAIR per rafforzare la provenienza dei dati nel flusso di lavoro dell'esperimento scientifico NFFA-DI, dalla creazione dei dati all'utilizzo degli stessi. L'insieme delle procedure sarà sviluppato

⁵ punto 4.2.2 della *Politica in materia dei Dati di Ricerca NFFA-DI*

⁶ punto 4.3.2 della *Politica in materia dei Dati di Ricerca NFFA-DI*

⁷ punto 4.3.5 della *Politica in materia dei Dati di Ricerca NFFA-DI*

⁸ punto 4.1.4 della *Politica in materia dei Dati di Ricerca NFFA-DI*

tenendo conto delle esigenze provenienti dalle varie comunità all'interno di NFFA-DI. Verrà prestata particolare attenzione all'identificazione e alla personalizzazione delle soluzioni esistenti di quaderni elettronici di laboratorio (ELN) e di sistemi di gestione delle informazioni di laboratorio (LIMS) per descrivere i flussi di lavoro e l'elaborazione dei campioni. Ci si focalizzerà sulla registrazione (semi-) automatizzata dei metadati durante gli esperimenti come passaggi iniziali per l'implementazione dei set di dati FAIR-by-design.

3.1.2. *Accessibility: rendere i dati accessibili*

Nel progetto NFFA-DI, i dati di ricerca e i metadati ad essi associati processati, analizzati e che in qualche modo siano legati a dei risultati pubblicati, saranno resi disponibili ad accesso aperto. Questo avverrà auspicabilmente il prima possibile o almeno entro la data di pubblicazione del relativo articolo scientifico o poster, facendo uso di una data repository aperta⁹. Il team di data management al momento ha individuato come NOMAD (<https://nomad-lab.eu/nomad-lab>) la repository d'elezione.

I dati e metadati raccolti nel sistema OFED e in particolare nella piattaforma NOMAD Oasis centrale, invece, resteranno accessibili con accesso ristretto agli autori degli stessi per dieci anni.

3.1.3. *Interoperability: rendere i dati interoperabili*

Per rendere i dati interoperabili, ovvero consentire lo scambio e l'integrazione dei dati tra ricercatori, istituzioni, organizzazioni, paesi, ecc., i dati di pubblicazione prodotti nel progetto NFFA-DI saranno in un formato tale da poter essere aperto con un software open source multipiattaforma¹⁰. In questo modo si renderà possibile gratuitamente a qualsiasi utente l'accesso, l'estrazione, l'utilizzo, la riproduzione e la diffusione. Per migliorare l'interoperabilità, si consiglia ai Destinatari di utilizzare schemi di metadati standard e ratificati per descrivere i dati di ricerca (standard di metadati), come quelli elencati nella directory degli standard di metadati della Research Data Alliance (<http://rd-alliance.github.io/metadata-directory/>), sul sito web del Digital Curation Center (<http://www.dcc.ac.uk/resources/metadata-standards>) o su FAIRsharing.org (<https://fairsharing.org/standards>)

Uno degli obiettivi principali di NFFA-DI è la creazione di un sistema avanzato per la gestione dei dati e dei metadati e l'implementazione di schemi di metadati per alcune delle tecniche di misura presenti nel catalogo NFFA-DI per le quali non è disponibile uno standard di metadati comunemente accettato. Questi schemi di metadati mirano ad essere comunemente accettati dalla comunità scientifica pertinente.

Il formato di riferimento che proponiamo per la creazione della nomenclatura, del vocabolario e degli schemi di metadati è quello sviluppato da FAIRmat nell'ambito di NOMAD o in alternativa NeXus (<https://www.nexusformat.org/>), opportunamente arricchito con nuove voci coerenti con esso.

I *Metadata Schema* da implementare dovranno essere in grado di descrivere, con una struttura e parametri standardizzati, set di dati ottenuti da varie tipologie di esperimenti, con l'obiettivo di diventare uno standard. Un insieme minimo di parametri obbligatori e consigliati sarà accompagnato da un dizionario contenente il più ampio insieme di Metadati che potrebbero essere necessari a seconda dell'Esperimento e della Strumentazione utilizzata, da aggiungere secondo necessità.

⁹ punto 4.3.2 della *Politica in materia dei Dati di Ricerca NFFA-DI*

¹⁰ punto 4.2.4 della *Politica in materia dei Dati di Ricerca NFFA-DI*

I fornitori e gli esperti coinvolti in NFFA-DI sono invitati a collaborare e a presentare i loro suggerimenti, eventuali standard di metadati già in uso per i loro strumenti scientifici e le loro idee per l'implementazione di nuovi schemi.

3.1.4. *Reusability: aumentare il riutilizzo dei dati*

Per consentire il più ampio riutilizzo possibile, tutti i dati e i metadati legati a dei risultati pubblicati verranno resi disponibili ad accesso aperto utilizzando una licenza CC-BY (Creative Commons Attribuzione), o una licenza con diritti equivalenti¹¹.

I documenti rilevanti, come i risultati finali o i report pubblici, saranno pubblicati su repository aperte e certificate come Zenodo e auspicabilmente saranno accessibili attraverso il sito del progetto.

In conformità con le linee guida FAIR, l'intento è che i dati di ricerca generati all'interno dell'infrastruttura possano essere riutilizzabili anche dopo la fine del progetto, se consentito, e supportato dalle risorse del progetto e dall'infrastruttura necessaria.

L'adozione di un approccio FAIR-by-design all'interno di tutti i laboratori afferenti al progetto permetterà di avere pieno controllo della provenienza dei dati, permettendo oltre a un riutilizzo efficiente degli stessi la riproducibilità degli esperimenti qualora ne sorgesse la necessità.

3.2. FAIR DATA MANGEMENT TEAM

Il progetto NFFA-DI prevede la raccolta dei dati in modalità FAIR-by-design, che verrà realizzata seguendo il workflow descritto nella sezione 2.3. Per l'attuazione del FAIR-by-design in tutti i laboratori che fanno parte del progetto è richiesta una forte collaborazione fra i PI dei singoli laboratori, chi gestisce OFED e i data manager e curator.

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 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 descritto nel paragrafo 2.2.

Per supportare e accompagnare i vari laboratori (vedi tabella 4) verso una gestione FAIR-by-design dei dati è stato istituito un FAIR DATA MANAGEMENT TEAM, di cui sono dettagliati i componenti nella tabella 6, e che potrà essere contattato per ogni esigenza relativa alla gestione FAIR dei dati, come ad esempio la definizione delle ontologie dei metadati e l'installazione delle Oasis di NOMAD.

FAIR DATA MANAGEMENT TEAM (FDTM)	
<i>Nome e Cognome</i>	<i>Istituto di appartenenza</i>
Ruggero Lot	Area Science Park
Federica Bazzocchi	Area Science Park

¹¹ punto 4.3.4 della *Politica in materia dei Dati di Ricerca NFFA-DI*

Mariarita de Luca	Area Science Park
Dario De Angelis	CNR-IOM
Irene Modolo	CNR-IOM

Tabella 6: composizione del FAIR DATA MANAGEMENT TEAM.

Una ulteriore attività di supporto proposta alle Unità UO è legata alla formazione nel campo della gestione FAIR dei dati. All'interno del WP8 è previsto un corso di perfezionamento intitolato **“Scientific Data Management and Curation”** con l'obiettivo di formare dottorandi o giovani ricercatori appartenenti ai vari partner di NFFA-DI sui temi legati ai principi FAIR e in particolare all'acquisizione dei dati in modalità FAIR-by-design nei laboratori afferenti al progetto, Questi studenti verranno supportati dal FAIR DATA MANAGEMENT TEAM che fungerà da punto di riferimento per la cura e la gestione dei dati all'interno del progetto NFFA-DI. Questa comunità eterogenea di data curators sarà incaricata di rivedere, rinforzare, pulire e standardizzare i dati e i relativi metadati assicurando la FAIRness degli stessi.

4. ALTRI PRODOTTI DELLA RICERCA

Il progetto NFFA-DI prevede la realizzazione di un'infrastruttura di ricerca che coinvolge svariati laboratori sul territorio italiano e questi dovranno acquisire i dati in modalità FAIR-by-design, dati che poi verranno raccolti su un Data Lake comune.

La realizzazione di un progetto così ambizioso potrà essere raggiunta sviluppando una serie di pipeline e tool ad hoc nonché software e interfacce personalizzabili. Tutti questi strumenti sono prodotti aggiuntivi del progetto stesso e potranno essere esempi di best practice e diffusi ad accesso aperto seguendo i principi FAIR.

Inoltre, un nodo del progetto si occupa di realizzare sensori micrometrici con vari spessori, depositando materiale apposito su piastre di silicio. In tal caso i prodotti della ricerca non sono solo i dati e i metadati dei processi che portano alla loro realizzazione, ma le strutture stesse, per le quali verrà studiato un sistema di catalogazione FAIR ad hoc.

5. ALLOCAZIONE DELLE RISORSE

I costi per lo sviluppo, l'infrastruttura e la gestione del sistema OFED saranno coperti dal finanziamento PNRR entro l'arco temporale del progetto e imputati a due specifici Work Packages (WP2 e WP3).

Inoltre, il CNR-IOM, coordinatore del progetto, ha firmato un accordo con AREA Science Park per ospitare tutti i servizi dati e le infrastrutture dati a disposizione di NFFA-DI all'interno di ORFEO, il data center di AREA Science Park.

Il WP3 avrà il ruolo di controllare e mantenere la corretta gestione dei dati, nonché di migliorarla quando si renderà necessario.

Le risorse per la conservazione a lungo termine non sono state ancora discusse, poiché ciò comporta il progresso generale del progetto e sarà affrontato in ulteriori versioni di NFFA-DI-DMP, tenendo presente che il progetto prevede di mantenere l'infrastruttura per 10 anni.

6. SICUREZZA DEI DATI

In generale, tutti i data center in cui sono archiviati i dati del progetto sono certificati in modo sufficiente a garantire la sicurezza richiesta, in particolare, per quanto riguarda la documentazione, le relazioni di progetto, e i dati personali. Tutti i servizi web atti all'interazione con questi servizi sono accessibili tramite il protocollo di trasferimento di ipertesto sicuro (https). I dati personali contenuti nei documenti archiviati non vengono né controllati né analizzati e sono sotto la responsabilità del creatore della risorsa. I documenti registrati sulla piattaforma sono conservati sui server del partner NFFA-DI responsabile del WP2 che offre gli standard di sicurezza informatica necessari a garantire la protezione da possibili fughe di dati. L'archivio sarà sottoposto a backup sia in loco che fuori sede per proteggere i dati da disastri, in questo modo l'archivio è protetto contro la perdita o il furto. Sarà implementata una politica di controllo degli accessi per consentire l'accesso fisico all'archivio. I dati personali saranno protetti da accessi non autorizzati, corruzione o furto per l'intero ciclo di vita del progetto.

I dati provenienti dalla ricerca saranno anch'essi serviti attraverso accessi sicuri (https e/o VPN) e verranno conservati sui server del partner AREA Science Park che offre, anche in questo caso, gli standard di sicurezza informatica necessari per garantire la protezione di tali dati.

6.1. Policy di Backup

Per quanto riguarda i Dati di Ricerca, sono previste al momento le seguenti policy di backup:

- i dati che saranno poi soggetti a pubblicazione saranno soggetti al backup che riguarderà l'intera piattaforma.
- i dati che risulteranno non soggetti a pubblicazione saranno salvati su File System distribuito dotato di erasure code 8+2, il failure domain è Datacenter. Per questi dati non è previsto backup, ma saranno predisposte diverse tipologie di ridondanza in base alla dimensione dei dati:
 - <20 GB: ridonati tra le UO e il Data Lake.
 - >=20 GB: non ridonati, a meno di diverse policy definite caso per caso tramite Lab-DMP.

Per i dati non pubblicati, si potrà valutare la possibilità di prevedere degli snapshot per il ripristino dei dati (in situazioni particolari), fino a 6 mesi.

Ulteriori dettagli relativi a questa sezione verranno inseriti in una futura versione di questo documento, successiva al bimestre 9, momento in cui è prevista la sottomissione del Deliverable 3.2.

I dati classificati come documentazione e relazioni di progetto descritti nella tabella 2 sono in carico al WP2. E' plausibile che il backup verrà effettuato ogni giorno; per evitare inconsistenze nei backup, durante l'esecuzione degli stessi, è previsto un periodo di downtime della piattaforma NFFA-DI, di durata dipendente dalla mole di dati di cui effettuare il backup. La conservazione dei dati di backup viene garantita per 2 anni, con la seguente granularità:

- Fino a 7 giorni: Giornaliera,
- Fino a 8 settimane: Settimanale,
- Fino a 6 mesi: Bimestrale,
- Fino a 2 anni: Semestrale.

6.2. Conservazione dei dati

Le tempistiche della conservazione dei Dati di Ricerca sono le seguenti:

- Dati soggetti a Pubblicazione: 10 anni (minimo legale).

- Dati non soggetti a Pubblicazione: attualmente la proposta è di tenerli per 5 anni di default, se non specificato diversamente nel Lab-DMP e concordato con il FDMT definito in tabella 6.

7. ASPETTI ETICI

I dati e metadati generati dalla strumentazione acquistata, implementata o integrata nel contesto del progetto NFFA-DI non includono questionari che trattano dati personali né sollevano questioni etiche per la loro condivisione con la comunità scientifica.

Tra i metadati che descrivono i dataset generati nei laboratori delle varie unità operative possono essere presenti dati personali dell'autore del dataset, allo scopo di riconoscerne l'attribuzione e garantire una corretta citazione e riconoscimento del lavoro di ricerca. Sugeriamo che le istituzioni partecipanti utilizzino formati di metadati che contengano il minimo numero di dati personali necessario a riconoscere l'attribuzione dell'insieme di dati, ma non abbiamo un controllo completo sulle pratiche di ciascun laboratorio. Questi dati personali vengono deliberatamente forniti dal ricercatore che genera i dati di ricerca, che può gestirne l'accessibilità (privata o pubblica). I dati personali che raccomandiamo di raccogliere sono: nome e cognome, ruolo, affiliazione, email e ORCID.

Ad ogni modo, tutti i dati personali raccolti dall'infrastruttura NFFA-DI nelle sue varie fasi operative rispetteranno i principi di limitazione della finalità, minimizzazione dei dati, accuratezza, limitazione della conservazione, integrità e riservatezza.

Tutti i dati personali vengono conservati con lo scopo di erogare il servizio e il loro trattamento sarà conforme al Regolamento (UE) 2016/679 del Parlamento Europeo e del Consiglio del 27 aprile 2016 - Regolamento generale sulla protezione dei dati (GDPR).

Il trattamento dei dati personali degli utenti che si registrano al sito www.nffa-di.it è regolato dalla Privacy Policy (<https://nffa-di.it/privacy-policy/>).

8. ANNEX 1: POLITICA IN MATERIA DEI DATI DI RICERCA NFFA-DI

1. Principi generali

1. Questa politica si applica ai Dati di Ricerca generati dalla strumentazione acquistata, implementata o integrata nel contesto del progetto NFFA-DI (dati grezzi o dati analizzati), inclusi i metadati associati.
2. L'infrastruttura NFFA-DI è tenuta a permettere accesso strutturato ed aperto ai dati scientifici e tecnologici, per amplificarne l'impatto ed il ritorno sulla società e sull'economia, sia a livello globale che a livello locale, come all'articolo 8 del Disciplinare del progetto (Allegato D).

2. Obiettivi

1. La presente politica mira a garantire che i Dati di Ricerca generati nel progetto NFFA-DI siano gestiti e resi accessibili per l'uso e il riutilizzo secondo gli obiettivi del Piano Nazionale Scienza Aperta e dell'European Open Science Cloud.
2. La presente politica mira a garantire la disponibilità a lungo termine di dati scientifici e tecnologici di valore per la ricerca, l'istruzione e per un più ampio sfruttamento da parte di individui, governi, imprese o altre organizzazioni.
3. La presente politica mira ad articolare le responsabilità di tutti gli utilizzatori della strumentazione acquistata, implementata o integrata nel contesto del progetto NFFA-DI rispetto all'archiviazione e alla conservazione dei dati di ricerca e dei metadati associati e derivanti dalle loro attività di ricerca.

4. La presente politica mira a fornire indicazioni sulle buone pratiche esistenti per la produzione e gestione di dati secondo i principi *FAIR* (reperibilità, accessibilità, interoperabilità e riusabilità).
3. Requisiti
 1. Gestione dei Dati di Ricerca
 1. Ciascun Responsabile di Laboratorio deve fornire, per ogni strumento scientifico acquistato, implementato o integrato nel contesto del progetto NFFA-DI, le informazioni sulla modalità di gestione dei Dati della Ricerca generati, utilizzando lo strumento messo a disposizione sul portale.
 2. Ciascun Principal Investigator a cui sia stato approvato un Proposal sull'infrastruttura NFFA-DI deve sottomettere un Data Management Plan (Proposal-DMP) relativo alla gestione dei i dati di ricerca prodotti, utilizzando lo strumento messo a disposizione sul portale NFFA-DI.
 3. La prima versione del Proposal-DMP deve essere sottomessa prima della data di inizio degli accessi programmati presso l'infrastruttura NFFA-DI per quello specifico Proposal, mentre la versione definitiva andrà sottomessa entro e non oltre un mese dalla data di conclusione degli esperimenti.
 4. Tutti gli utilizzatori della strumentazione acquistata, implementata o integrata nel contesto del progetto NFFA-DI sono tenuti ad utilizzare il quaderno elettronico di laboratorio integrato nel sistema OFED, al fine di facilitare buone pratiche di gestione dei dati, condivisione di dati e documentazione tra ricercatori, dimostrare la provenienza e proteggere dalla perdita di dati.
 2. Principi *FAIR*
 1. Tutti i dati di ricerca e metadati associati generati dalla strumentazione acquistata, implementata o integrata nel contesto del progetto NFFA-DI devono essere gestiti secondo i principi *FAIR*, con particolare cura nel caso siano presenti o discussi in Pubblicazioni Scientifiche, presentazioni a conferenze e brevetti.
 2. I dati di ricerca generati dalla strumentazione acquistata, implementata o integrata nel contesto del progetto NFFA-DI devono essere archiviati sul sistema OFED con permessi di accesso specifici per ciascun utente. OFED ne garantisce lo stoccaggio sicuro, permette di assegnare loro un identificatore univoco universale (UUID) e fornisce un'appropriata integrazione di dati e relativi metadati.
 3. Se esistente, si raccomanda l'uso di uno standard di metadati formale. Una panoramica degli standard di metadati esistenti organizzati per disciplina può essere trovata sui siti web della Research Data Alliance (<http://rd-alliance.github.io/metadata-directory/>), del Digital Curation Center (<http://www.dcc.ac.uk/resources/metadata-standards>) o FAIRsharing.org (<https://fairsharing.org/standards>). NFFA-DI consiglia di utilizzare il linguaggio NOMAD sviluppato nell'ambito del progetto FAIRmat o il formato NeXus.
 4. I dati di ricerca generati dalla strumentazione acquistata, implementata o integrata nel contesto del progetto NFFA-DI devono essere archiviati in un formato di file che possa essere aperto con un software multiplatforma open source.
 3. Accesso aperto
 1. Gli utilizzatori della strumentazione acquistata, implementata o integrata nel contesto del progetto NFFA-DI devono esaminare la possibilità di tutelare i propri risultati nel caso in cui si preveda ragionevolmente che questi possano essere sfruttati commercialmente o industrialmente.
 2. I dati di ricerca e metadati associati generati dalla strumentazione acquistata, implementata o integrata nel contesto del progetto NFFA-DI presenti o

discussi in pubblicazioni scientifiche e presentazioni a conferenze (dati grezzi e dati analizzati) devono essere resi disponibili ad accesso aperto entro la data di diffusione del prodotto, utilizzando una data repository aperta, in grado di assegnare loro un Persistent Identifier (PID).

3. I metadati dei dati resi disponibili ad accesso aperto devono includere informazioni sugli strumenti necessari a riutilizzare o convalidare i dati (ad esempio software o codici specifici, algoritmi e protocolli di analisi)
4. I metadati dei dati resi disponibili ad accesso aperto devono essere distribuiti sotto licenza CC-BY (Creative Commons Attribuzione) o altra licenza con diritti equivalenti.
5. I destinatari devono identificare i dati depositati indicando nella deposizione almeno i seguenti metadati amministrativi:
 - Breve descrizione, Data di deposito e Autore(i) del Dataset
 - NextGenerationEU - PNRR "NFFA-DI" IR0000015 - 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"
 - Nome, acronimo e numero del progetto di sovvenzione: "NANO FOUNDRIES AND FINE ANALYSIS - DIGITAL INFRASTRUCTURE, NFFA-DI, IR0000015"
 - Termini di licenza: CC BY o una licenza con diritti equivalenti
 - Numero ID del proposal (se applicabile)
 - PID delle pubblicazioni correlate e di altri prodotti di ricerca (se applicabile).

9. ANNEX 2: GLOSSARIO

Data Lake: Un archivio di dati che consente di memorizzare grandi quantità di dati grezzi, strutturati e non strutturati, in un unico repository centralizzato. La caratteristica distintiva di un data lake è la sua capacità di accogliere dati in vari formati, senza richiedere una predefinita struttura o schema. Le principali caratteristiche di un Data Lake includono: scalabilità, flessibilità, diversità dei dati, accesso unificato ed elaborazione dei dati (spesso, i data lake vengono utilizzati in combinazione con strumenti di elaborazione dati).

Data Management Plan (DMP): Un documento formale che delinea come saranno gestiti i dati durante e dopo un progetto di ricerca. Descrive il tipo di dati che verranno utilizzati per la ricerca, come questi dati vengono raccolti, organizzati e archiviati, e in quali formati. Specifica come i dati saranno accessibili e documentati per la condivisione e il riutilizzo durante e dopo la conclusione del progetto.

Data Repository: Un sistema informativo operativo per la gestione e l'organizzazione delle risorse digitali. Il Data Repository utilizza metodi di organizzazione strutturati, schemi standardizzati e metadati per garantire che i dati siano facili da trovare e permette di definire regole specifiche per accedervi. Un Data Repository può essere associato a una determinata istituzione o a un insieme di istituzioni, a uno strumento specifico o a un gruppo di strumenti, o può essere gestito da una terza parte. I Data Repository possono essere utilizzati direttamente dagli utenti di ricerca o meno.

Dati di Ricerca: Informazioni (in particolare fatti o numeri) raccolte per essere esaminate e prese in considerazione, al fine di servire da base per il ragionamento, la discussione o il calcolo in un contesto di ricerca. Esempi di Dati di Ricerca includono statistiche, risultati di esperimenti,

misurazioni, osservazioni derivanti da lavoro sul campo, risultati di sondaggi, registrazioni di interviste e immagini. I Dati Grezzi (Raw Data) e i Dati Analizzati (Analyzed Data) sono tipi particolari di Dati di Ricerca.

Dati Personali: Informazioni che identificano o rendono identificabile, direttamente o indirettamente, una persona fisica e che possono fornire informazioni sulle sue caratteristiche, le sue abitudini, il suo stile di vita, le sue relazioni personali, il suo stato di salute, la sua situazione economica, ecc. I Dati Personali trattati in questo documento sono i dati che permettono l'identificazione diretta, come i dati anagrafici, e i dati che permettono l'identificazione indiretta, come un numero di identificazione (ad esempio, il codice fiscale, l'indirizzo IP, il numero di targa).

Metadati/Metadata: Un insieme di informazioni descrittive, strutturali e contestuali che descrivono il contesto, il contenuto e la struttura dei Dati di Ricerca e/o dei Dataset e la loro gestione nel tempo. Forniscono informazioni relative a progetti di ricerca, inclusi (ma non limitati a) il contesto dell'esperimento, gli utenti di ricerca, i metodi di analisi dei dati e altre informazioni logistiche. I metadati possono includere descrizioni su come i dati e i file sono denominati, strutturati fisicamente e archiviati.

Metadata Schema: Un piano logico che mostra le relazioni tra i Metadati, normalmente attraverso la definizione di regole per il loro uso e gestione con specifico riguardo alla semantica, alla sintassi e alla cardinalità (obbligatoria, facoltativa, consigliata) dei valori. Uno schema di metadati che ottiene un'ampia accettazione da parte di una comunità di utenti di riferimento ed è stato formalmente approvato dalle organizzazioni di standardizzazione, diventa un Metadata Standard.

Metadata Standard: Un Metadata Schema che soddisfa le esigenze di una comunità scientifica ha ottenuto il consenso ed è stato ratificato come standard da alcuni organismi ufficiali. Un Metadata Standard descrive le informazioni e i termini necessari per definire correttamente dati specifici e favorisce l'interoperabilità.

Persistent Identifier (PID): Riferimento che punta a una risorsa digitale online oppure relativa a qualcosa o qualcuno nel mondo reale. I PID sono progettati per rimanere univoci e inalterati a tempo indefinito, anche se la posizione della risorsa su internet o nel mondo reale varia. I PID rappresentano una parte essenziale del processo di adozione del paradigma dell'Open Science perché consentono una affidabile e immediata identificazione di elementi connessi ai processi dell'attività scientifica. Esempi di PID sono: DOI e ORCID.

Principal Investigator (PI): Persona responsabile dello svolgimento di un esperimento, a capo del team di un Proposal.

Proposal: Richiesta di accesso alle infrastrutture di ricerca offerte da NFFA-DI attraverso il sistema implementato sul sito web del progetto al fine di svolgere attività di ricerca.

Pubblicazione Scientifica: Qualsiasi delle seguenti contribuzioni (sottoposte a peer-review o meno): articolo in una rivista scientifica (e relative informazioni di supporto), monografia, libro o capitolo di libro, atti di conferenza e 'letteratura grigia' (cioè materiale pubblicato informalmente e non sottoposto a un processo editoriale standard, ad esempio rapporti e sintesi).

Responsabile di Laboratorio: La persona che ha la responsabilità generale delle operazioni tecniche, scientifiche e amministrative del laboratorio/gruppo di ricerca. Essa è responsabile di garantire che il laboratorio sia conforme agli obblighi e alle normative UE in merito ai dati di ricerca. Solitamente, questa persona è qualcuno che ha esperienza con una o più delle tecniche di misura offerte e dovrebbe avere una conoscenza chiara dei dati prodotti e di come questi dati vengono gestiti durante il processo di ricerca. Ogni laboratorio può avere uno o più Responsabili di Laboratorio.