WHITE PAPER · KEY INSIGHTS

Non-technical barriers to data management

in EU materials characterisation and modelling projects

Diagnosis, recommendations, and a research agenda for FP10

Alexandre Ouzia (HeidelbergMaterials) and Alex Simperler (EMMC)
Heidelberg Materials · MatCHMaker
MatCHMaker Exploitation Workshop · EMMC

INTRODUCTION

Why this paper, and why now

April 2025 — Vienna
Six HE sister projects + EMMC mapped data-lifecycle pain points and good practice. The output was descriptive and constructive (Ouzia et al., 2025).

Feb–Apr 2026 — Interviews & Survey
Eight semi-structured, hour-long interviews with senior practitioners across the same network.

May 2026 — White paper
Diagnoses non-technical barriers, places them against the literature, and proposes recommendations for FP10.

The thesis

Technical infrastructure for FAIR data — ontologies, repositories, FAIR principles, automated extraction tools — has matured **but is not adopted**. The binding constraint is no longer the technology. It is **the human, organisational and policy environment** in which the technology is asked to operate.

METHODOLOGY

What this paper rests on

8	12	9	ADKAR
semi-structured interviews	corroborative survey responses	EU projects covered (6 core + 3 referenced)	diagnostic framework

Sample: WP leads, data-pillar leads, coordinators across MatCHMaker, KNOWSKITE-X, AID4GREENEST, D-STANDART, CoBRAIN, AddMorePower; PINK, COMPASS, OntoTrans referenced through interviewees with leadership in those projects.

Instrument: two-half guide — diagnostic (ADKAR) + constructive (17-item rapid grid + open scenarios). All interviews under Chatham House Rule.

Caveats: n = 8 supports thematic claims, not quantitative ones. Sample biased toward characterisation-side and toward partners with explicit data-pillar responsibility — by design. No independent document review of DMPs or deliverables.

FINDINGS · OVERVIEW

Eight cross-cutting findings

1. **Time is the master barrier**
   Time pressure dominates every other constraint.

2. **DMP as a compliance artefact**
   Filed for the funder, not used for coordination.

3. **Projects rise or fall on “double-hat” people**
   Hybrid science + data engineering is the binding constraint.

4. **Subculture, not sector**
   Modeller vs experimentalist, not academia vs industry.

5. **DM value accrues after the project ends**
   Cost is borne by data producer; benefit by future user.

6. **EU reporting cascade is too slow**
   Three periods over four years cannot shape behaviour.

7. **Ontologies arrive too late**
   Models mature on a 5–20 year cycle; projects last 3–4.

8. **Lump-sum weakens cross-partner data work**
   Cross-partner tasks are the first to fragment.

FINDINGS 1 & 2

Time and the DMP-as-compliance-artefact

F1 — Time is the master barrier

Named in every interview as the dominant operational constraint.

Form varies by sector — publication pressure, commercial deliverables, multi-project commitments — but the destination is the same.

DM consistently loses the daily competition for hours, even where awareness and willingness are present.

F2 — DMP as compliance artefact

Interviewees described their DMPs as “filed”, “somewhat followed”, even “effectively meaningless”.

The pattern is not absence — it is that DMPs are written for the funder, not for internal coordination.

THE COHORT AUDIT (May 2026)

Of the six HE sister projects funded under the same 2022 call topic: **one** publishes a downloadable DMP; **one** maintains an active dataset community on a registered repository; **one** operates an active public source-code organisation. Cohort mean accessibility score: **2.2 / 5**, with a tenfold variation between strongest and weakest despite identical funding scheme and similar budgets.

FINDINGS 3 & 4

Double-hat people, and why subculture beats sector

F3 — Double-hat individuals

Multiple interviewees independently named the same workforce phenomenon: a project's DM stands or falls on one or two individuals fluent in both the science and the data engineering.

These people — “shepherds”, “double-hatted”, “the one PhD who came from industry” — bridge communities that otherwise communicate poorly.

They are scarce, costly to themselves (overload, role ambiguity), and unrecognised by the reward system.

F4 — Subculture, not sector

The instinctive framing — academia bad at DM, industry good — is not what we observed.

In one interview a corporate test house produced the least usable metadata; in another the academic partners were the bottleneck while industry was already disciplined.

What does correlate is professional subculture — the tacit conventions, file structures and acronyms a lab accumulates over years. Asking partners to change is asking them to override years of practice, regardless of sector.

FINDINGS 5 & 6

Deferred value, and a reporting cadence built for a different project

F5 — DM value accrues after the project ends

Within the 3–4 year project window, the people who hold the contextual knowledge are present.

The pain of poorly-documented data is felt later, by future users, after those people have left.

This is a collective-action problem, not a failure of individual virtue. Awareness campaigns alone cannot close it.

F6 — Reporting cascade is too slow

Horizon Europe: typically three reporting periods over four years. First review at month 18.

Quarterly or sprint cadences — used in Fraunhofer-internal, German national, some industry projects — produce smaller, more frequent commitments that distributed teams observably honour.

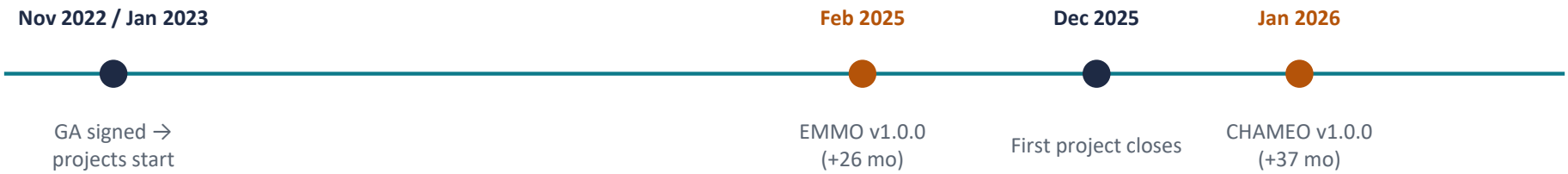
The mismatch between EU cadence and the iterative reality of R&D is a non-trivial contributor to the late arrival of data infrastructure (see F7).

FINDINGS 7 & 8

Late ontologies, and lump-sum's effect on data work

F7 — Ontology readiness arrives too late

All six sister projects committed at GA signature (Nov 2022) to deliverables involving EMMO, CHADA, MODA, CHAMEO. The standards themselves matured during the projects, not before them.



F8 — Lump-sum funding may weaken cross-partner data work

Hypothesis (early-wave LS cohort still young, but consistent with principal-agent theory): lump-sum allocation, by tying payment to WP-level technical completion, raises incentive intensity on the well-measured task and starves cross-partner data work — which is measured only through DMP narrative.

DIAGNOSIS

Where the change actually fails: an ADKAR reading

A	**Awareness**	Partial	High in DM leads & ontologists. Lower among PIs and PhD students.
D	**Desire**	Failing	The dominant ADKAR failure. Cost borne by producer; benefit by future user.
K	**Knowledge**	Inert	Adequate in supply, weak in deployment: one-shot, generic, voluntary.
A	**Ability**	Constrained	Tools mixed; time severely constrained. F1 lives here.
R	**Reinforcement**	Almost absent	Mid-term reviews check existence of DMP, not implementation. No sanctions. No rewards.

ADKAR is empirically validated in healthcare and education, not in research consortia. Useful diagnostically; agnostic on systemic drivers — exactly the level at which F1, F2, F5, F6, F8 operate.

The ADKAR change-management framework, mapped to MCM data-management practice

Each finding sits at the stage where it primarily obstructs adoption

FINDINGS THAT BITE HERE

- F5 Deferred value

FINDINGS THAT BITE HERE

- F1 Time pressure
- F5 Deferred value
- F4 Subculture inertia

FINDINGS THAT BITE HERE

- F7 Ontologies arrive late

FINDINGS THAT BITE HERE

- F1 Time pressure
- F3 Double-hat scarcity
- F8 Source-attached metadata

FINDINGS THAT BITE HERE

- F2 DMP as compliance
- F6 Reporting cadence too slow

Interventions placed only at Awareness or Knowledge cannot fix a problem whose root sits at Desire, Ability or Reinforcement.

Adapted from Hiatt (2006).

RECOMMENDATIONS · THE SPINE

Five concrete reform proposals for FP10

1

DM steward budget line

Mandated, audited, sized 1.5–3% of project direct costs. Evidence-based ratio (commission the workforce study).

2

Mid-term DMP implementation review

Certify implementation and use, not existence. ~€10–25k/project.

3

Metadata-model anchoring at proposal stage

Adopt-and-extend an existing model, or fund the new model as first-class deliverable with sequencing implication.

4

Project-level data maturity badge

peer-assessed. Crowds intrinsic motivation in; fines crowd it out.

5

Scaled EU DM-as-shared-service

Toward per-capita parity with NFDI by mid-FP10. Domain consortia with multi-year commitments, not project-cycle.

Multi-task moral hazard reading: *the steward line, the badge and the maDMP review jointly upgrade measurability of the noisy task — exactly what theory predicts is needed to neutralise lump-sum's incentive distortion.*

THE GATE MODEL IN PRACTICE · MATERIALS CHARACTERISATION & MODELLING

An ideal FP10 cement project, month 0 to 48

One shared, living data space is the spine. Source-attached capture feeds it; a supported gate recurs before every reporting period. **Operationalises Recs 1–3.**

THE CEMENT WORKSTREAM · METADATA EMITTED AT SOURCE

CHARACTERISATION

- **XRD** 2θ range · step · anode
- **SEM** kV · WD · detector
- **Mortar strength** age · w/c · load rate

MODELLING

- **ML strength model** train set · features · version
- **Finite-element** mesh · model · BCs

PROJECT LIFECYCLE · THE SHARED DATA SPACE PERSISTS THROUGHOUT

M0 — **Kick-off**
partners seed the space with one of every file type
F8

≈ M15 — **Gate → RP1**
1:1 steward sessions in the 4 wks before review
F6 F1

≈ M33 — **Gate → RP2**
model + descriptions iterate on real data

M48 — **Gate → RP3 · close**
production-grade, FAIR, reusable deposit
F2

A **Ability — scheduled time, steward & source tools**

R **Reinforcement — a gate before every review**

Inside one DM gate: four sessions, four outputs

Four weekly one-hour sessions per data producer, three months before each review. The work-package lead asks the domain questions; the EC-provided steward asks the structuring ones.

THE FOUR-WEEK WINDOW

4 × 1 hour

per data producer

F1 *scheduled, supported hours*

- **Data steward** *structuring Qs* — Rec 1
- **Data producer** *brings real data*
- **WP lead** *domain Qs*

1. **Updated metadata** — fitted to the consortium working model — Rec 2
2. **Cross-partner fixes** — column names · sample IDs · units — F2
3. **Model extension** — where the working model can't yet hold it
4. **Deliverable material** — figures, tables, methods text — F5

THE VALUE LOOP

Output 4: data described & drafted

↓

Feeds the partner's OWN deliverable

benefit returns in the SAME reporting cycle

D **Desire — producer sees the payoff**

Why it works: the gate replaces an unsupported mandate with scheduled support (Ability), its recurrence supplies the missing reinforcement, and closing the loop in-cycle repairs the producer-vs-future-user split that Section 5 names as the dominant Desire failure.

INTERNATIONAL COMPARATORS

What national systems already do — and the EU gap

GERMANY

NFDI

≈ €90 M / yr

27 domain consortia

Cautiously positive 2025 evaluation; recommendations on governance & sustainable financing.

UNITED KINGDOM

UKRI DRI

≈ £1.1 bn / yr

Compute Roadmap (2025)

Sequenced AI, exascale and data investments at scale.

UNITED STATES

NIH DMS Policy

Direct costs allowed

Effective Jan 2023

DM costs explicit in budget. Institutional cost ≈ $1M/yr at mid-large institutions (COGR).

EUROPEAN UNION

EOSC + Skills4EOSC + FAIR-IMPACT

≈ €32M + €6.5M + €8M

Across the entire stack

Roughly half of NFDI per researcher; two orders below UKRI.

Where the change actually has to happen

We resist the simpler framing that the European Commission alone holds the lever.

European Commission

Sets the terms — steward line, mid-term review, model anchoring, badge, shared-service scaling, lump-sum evaluation.

Research-performing organisations

Recognise data work in promotion, hiring and proposal evaluation.

Project coordinators

Run the consortium meeting differently. Quarterly cadence. Embed the practice. Use the badge to invite, not the fine to repel.

Individual researchers

Do the work — when the system around them recognises it and when they see the personal benefit. F5 (deferred value) is structural; closing it is collective.

Each level is necessary; none is sufficient. FP10 has the chance to be the first European framework programme designed around what is actually known.

alexandre.ouzia@heidelbergmaterials.com · MatCHMaker × EMMC · A. Ouzia & A. Simperler

Thank you for your attention !

Acknowledgements

The eight interviewees

Whose candour, under Chatham House Rule, is the empirical asset of this paper.

The MatCHMaker × EMMC sister-project network

MatCHMaker, KNOWSKITE-X, AID4GREENEST, D-STANDART, CoBRAIN, AddMorePower — and PINK, COMPASS, OntoTrans through their participants. The April 2025 Vienna workshop laid the empirical foundation.

Funding

MatCHMaker (GA 101091687); AddMorePower (GA 101091621); AID4GREENEST (GA 101091912); CoBRAIN (GA 101092211); KNOWSKITE-X (GA 101091534); D-STANDART (GA 101091409). Horizon Europe, European Union.

Thank you · alexandre.ouzia@heidelbergmaterials.com

B A C K U P

Material for Q&A

Anticipated tough questions: lump-sum politics · the 1:20 ratio defence · AI in DMPs · NFDI parity cost · why the cohort didn't do better · industry IP · references

The ADKAR change-management framework, mapped to MCM data-management practice

Each finding sits at the stage where it primarily obstructs adoption

FINDINGS THAT BITE HERE

- F5 Deferred value

FINDINGS THAT BITE HERE

- F1 Time pressure
- F5 Deferred value
- F4 Subculture inertia

FINDINGS THAT BITE HERE

- F7 Ontologies arrive late

FINDINGS THAT BITE HERE

- F1 Time pressure
- F3 Double-hat scarcity
- F8 Source-attached metadata

FINDINGS THAT BITE HERE

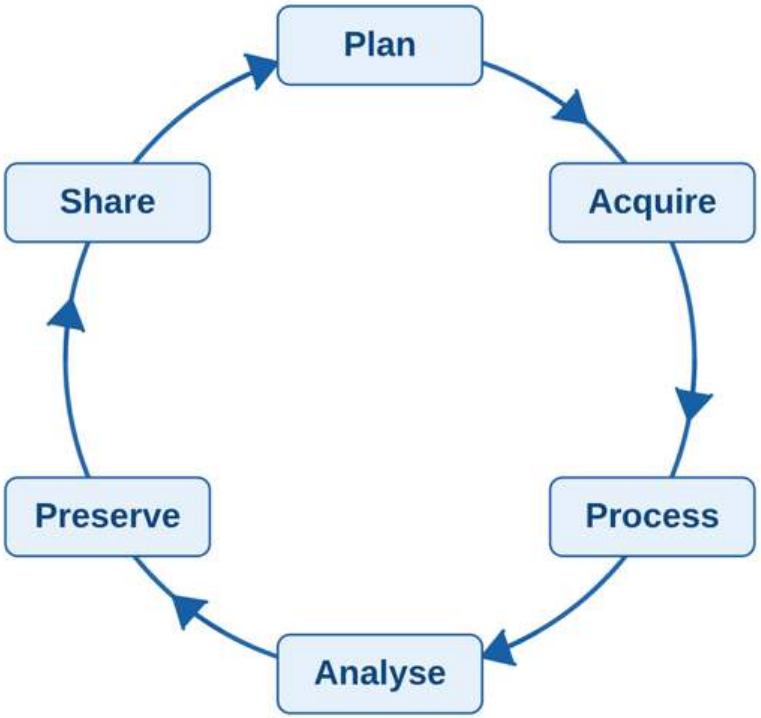
- F2 DMP as compliance
- F6 Reporting cadence too slow

Interventions placed only at Awareness or Knowledge cannot fix a problem whose root sits at Desire, Ability or Reinforcement.

Adapted from Hiatt (2006).

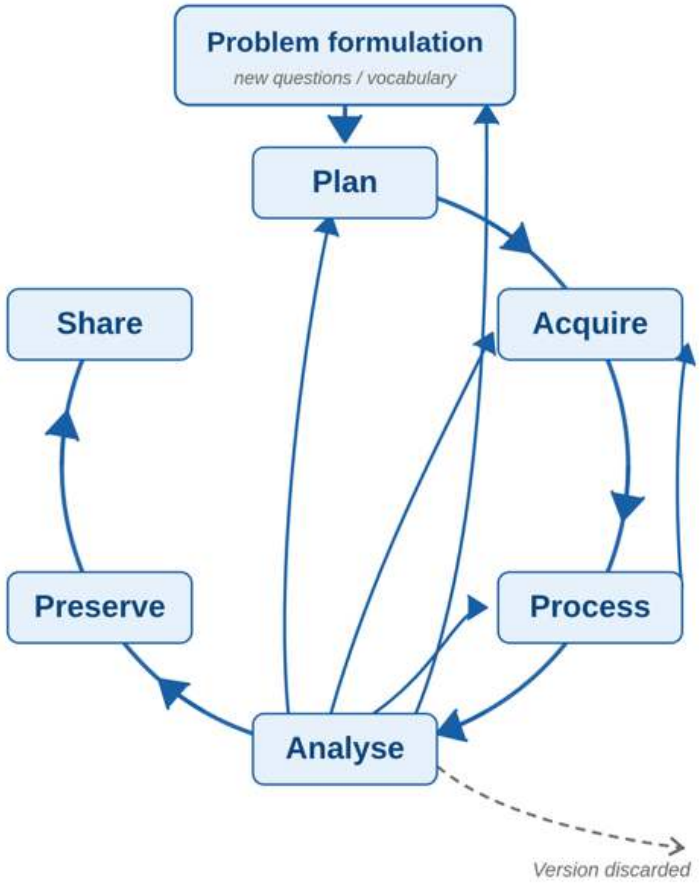
Idealised lifecycle

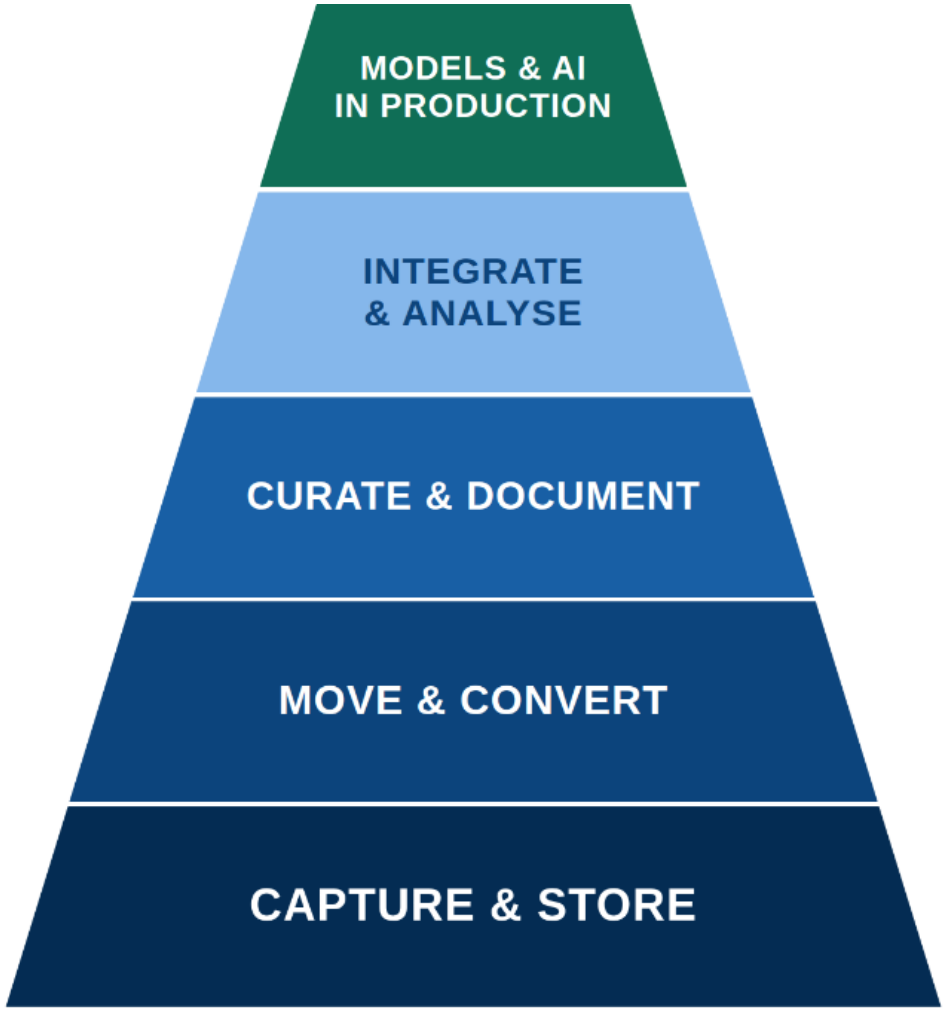
As drawn at the start of the workshop



Exploratory project reality

Iteration, reformulation, discarded paths





Not just a script that runs once, but a tool used by other people on new data.

- Cross-partner data integration
- Reusable analysis pipelines
- Shared, reproducible workflows

- Curated metadata (CHADA / MODA standards)
- Provenance and lineage
- Documented experimental context

- Conversion between proprietary formats
- Movement between partners
- Partner-to-partner data agreements

- Source-attached metadata (auto-captured)
- Agreed file-naming conventions
- Reliable storage

Without this layer, nothing above is usable.

"You cannot operate at the apex while the base is broken."

DISCUSSION · AI/LLMS

AI in research DM: mature, pilot, or watch

Capability	Status	Recommendation
LLM extraction from journal abstracts/sentences	*Mature*	**Use Now**
Domain transformers (MatBERT/MatSciBERT) for NER	*Mature*	**Use Now**
Interactive HITL image segmentation (bioimaging-grade)	*Mature*	**Use Now**
LLM extraction from MSE tables, figures, SI	*Immature*	**Pilot**
LLM-drafted DMPs	*Unevaluated*	**Pilot**
Ontology-grounded LLM KG population (CHAMEO/EMMO)	*Promising*	**Pilot**
LLM agents on cloud labs (ChemCrow / Coscientist class)	*Demonstrated*	**Monitor**
Foundation models for materials EM	*Absent*	**Monitor**

Caution. The most-cited controlled study reports ~18% fabricated citations for GPT-4 and ~55% for GPT-3.5 in multidisciplinary tests (Walters & Wilder, 2023). Retrieval grounding reduces but does not eliminate fabrication. *Principle: human-in-the-loop, with the human's time freed by AI rather than replaced. Prefer EuroHPC-trained open-weight models where performance is acceptable.*

Lump-sum: what we are saying — and what we are not

What we are saying.

Lump-sum funding, by tying payment to WP-level technical completion, raises incentive intensity on the well-measured task. Multi-task moral hazard predicts that this starves the noisy task — cross-partner DM. Beneficiary associations (EARTO, LERU, EUA, CESAER, the Guild) report fragmentation patterns consistent with this prediction.

What we are not saying.

We are not saying lump-sum should be reversed. The EC's 2024 assessment is positive on administrative-cost grounds, the political reality is that LS will reach ~50% of HE call budget by 2027, and almost no closed grants have completed a full lifecycle. We do not have evidence to recommend reversal — and would not recommend it on these data.

Our position.

Pause the further roll-out (the 2027 target) until an independent evaluation specifically including cross-partner data outcomes is in hand. In the meantime, install the guardrails — Recommendations 1, 2, 4 — that the multi-task model identifies as the right intervention.

The 1:20 data-steward ratio: where it comes from

Origin. Mons (2020), Nature 578:491 — “Invest 5% of research funds in ensuring data are reusable”. The exact phrasing: “I am convinced that doing so will require a large cadre of professionals, about one for every 20 researchers.” Expert judgement, plus a back-of-envelope calculation in the same article (PhDs lose ~80% of time to data munging; 20 stewards for 400 PhDs would recoup 200 FTE).
Prior literature. No prior published source uses 1:20. Not the EOSC HLEG report (2016), not Turning FAIR into Reality (EC 2018), not Mons's 2018 textbook.
Observed implementations. TU Delft (Andrews Mancilla et al. 2019): ~1:400–500 with embedded faculty stewards. Wageningen UR: ~1:333. Ghent University: ~1:1,500 (six FTE for ~9,000 researchers). Federated/hybrid models dominate among well-functioning services (RDA Professionalising Data Stewardship survey, 2022).
Our position. The ratio is folklore. FP10 should mandate steward provision and commission the workforce study to set evidence-based benchmarks in parallel.

BACKUP · B3

AI-drafted DMPs: why caution before adoption

The state of the evidence.

No peer-reviewed empirical comparison of LLM-drafted vs human-authored DMPs has been published. ARGOS, DMPonline and DMP Tool now expose hooks suitable for LLM integration; the OpenAIRE Research Graph can ingest ARGOS-produced maDMPs. But the canonical 2022 paper acknowledges that maDMP use cases “remain aspirational”.

Why this matters.

An LLM that drafts a fluent DMP without an evidence base for accuracy is exactly the tool that makes “benign simulation” worse. Hallucinated metadata, fabricated provenance and silent quality drift are documented failure modes (Walters & Wilder, 2023: 18% fabricated citations for GPT-4).

What to do instead.

Pilot AI-drafted DMPs in FP10 with controlled evaluation built into the project plan: measure researcher hours saved AND DMP quality vs human comparator. Until that evidence exists, treat AI as draft assistant under human authoring, not author.

NFDI parity for the EU: order-of-magnitude costing

NFDI baseline.

≈ €90 M / yr across 27 domain consortia, for ~80 M population in Germany.

EU per-capita parity.

Scaling NFDI's per-population spend to the EU's ~450 M would imply an envelope around €500 M / yr — clearly out of scope for a single FP10 cluster, and not what we are proposing.

Realistic FP10 target.

An envelope of €80–120 M / yr through FP10 would: (i) match NFDI in absolute scale, (ii) sit well below UKRI DRI's £1.1 bn, (iii) operate as a complement to national systems rather than a replacement. Delivered through domain-specific consortia with multi-year commitments, not project-cycle funding.

The order-of-magnitude argument.

PwC EU Services (2018): cost of inaction ~€10.2 bn / yr. Even a coarse 1% reduction repays the proposed envelope ~10× over.

An honest answer: why hasn't the cohort done better?

Three structural reasons.

1. Ontologies arrived too late. EMMO 1.0.0 in Feb 2025 was 26 months in; CHAMEO 1.0.0 in Jan 2026 was 37 months in. Three of six projects had key deliverables predate the standards they were supposed to use.

2. No funded “cluster CSA”. Coordination across the six sister projects has been informal — led by MatCHMaker hospitality and EMMC secretariat — without a budgeted instrument.

3. The data steward role is informal in nearly every project. F3 (the double-hat finding) plus the absence of an audited DM budget line means the work falls on whoever volunteers.

What the variation tells us.

The cohort's tenfold variation in observable open-science output, despite identical funding scheme and similar budgets, says the gap is closeable. AddMorePower's NOMAD parser stack and D-STANDART's CHADA-tagged Zenodo deposits show what is feasible inside the same constraints we all share.

Industry IP and openness: how the white paper handles it

What we observed.

Industry partners did not mechanically reduce open-science output in our cohort. The strongest public code stack belongs to a project with a major industry partner; the strongest public data community belongs to a project with industrial demonstrators. Conversely, the project with the strongest “silence” pattern has industry presence too. IP friction is plausible but does not by itself explain the variance — Finding 4 again.

The mechanism we recommend.

Embargoes, not exclusion. The MVDP (§7.3) and the maDMP review (Rec 2) make IP-affected datasets explicit at proposal stage with an embargo period and a public-release commitment. “Open by default, closed by exception, with named exception” rather than “closed by default”.

What this is not.

We do not recommend mandatory openness for IP-sensitive industrial data within the project window. Industry concerns are legitimate; the goal is to make them explicit and time-bounded, not to overrule them.

Selected references

Bowker & Star (1999). Sorting Things Out. MIT Press.

Bran et al. (2024). Augmenting LLMs with chemistry tools. Nat. Mach. Intell. 6, 525.

Del Nostro, Goldbeck & Toti (2022). CHAMEO. Applied Ontology 17(3), 401.

EARTO (2023). Members' experiences with lump-sums in HE.

European Commission, DG R&I (2024). Assessment of Lump Sum Funding in H2020 and HE.

Galison (1997). Image and Logic. U. Chicago Press.

Holmström & Milgrom (1991). Multitask Principal-Agent Analyses. JLEO 7, 24.

Knorr Cetina (1999). Epistemic Cultures. Harvard UP.

LERU (2025). Simplification in Horizon Europe.

Long, Cunningham & Braithwaite (2013). Bridges, brokers and boundary spanners. BMC HSR 13, 158.

Mason, Box & Burns (2020). PLOS ONE 15(8), e0238071.

Mons (2020). Invest 5% of research funds. Nature 578, 491.

Naranjo Bautista (2025). Tirole on EU research funding. Science|Business, 28 Aug.

Ouzia et al. (2025). Workshop on Data Lifecycles. Zenodo 16992583.

Piwowar & Vision (2013). Data reuse and the open data citation advantage. PeerJ 1, e175.

Polak & Morgan (2024). ChatExtract. Nat. Commun. 15, 1569.

PwC EU Services (2018). Cost of not having FAIR research data.

Schedler (2025). Horizon Europe's collaborative projects. THE, 24 Oct.

Scheffler et al. (2022). FAIR data for materials. Nature 604, 635.

Smale et al. (2020). DMP review. IJDC 15(1), 525.

Spichtinger (2022). DMPs in Horizon 2020. Open Research Europe 1, 42.

Stephan (2012). How Economics Shapes Science. Harvard UP.

Suhr, Dungl & Stocker (2020). MSE qualitative interview study. PLOS ONE 15(9), e0239216.

Tenopir et al. (2020). Data sharing worldwide. PLOS ONE 15(3), e0229003.

Walters & Wilder (2023). Fabricated citations from ChatGPT. Sci. Rep. 13, 14045.

Wenger (1998). Communities of Practice. Cambridge UP.

Wilkinson et al. (2016). FAIR Guiding Principles. Sci. Data 3, 160018.

Hiatt (2006). ADKAR. Prosci.

ECA Special Report 28/2018; Heitor (2024) interim evaluation HE.

White paper v0.2 (2026). Available from MatCHMaker / EMMC on request.