

Challenges in documenting health research data using the DDI standard

European DDI Users
conference 2025

3rd December 2025

Alessandro Morichetta (Inserm, France)

France Recherche en Santé Humaine (FReSH)

**Building a national catalogue
of individual data produced
by health research**

M
E
T
A
D
A
T
A

- Cohorts and other observational studies
- Clinical trials
- Registries
- Social science and humanities surveys in health

2023

Funding

"Mapping and qualifying
health databases"



end

2025

FReSH V1

Contributor interface

Migration and
standardisation of PEF
data

2027

FReSH V2

Harvesting **clinical
trials registries**

FReSH V3

Harvesting **social and
humanities science
catalogues, cohorts
and other data**

Accompanied by



Supported by



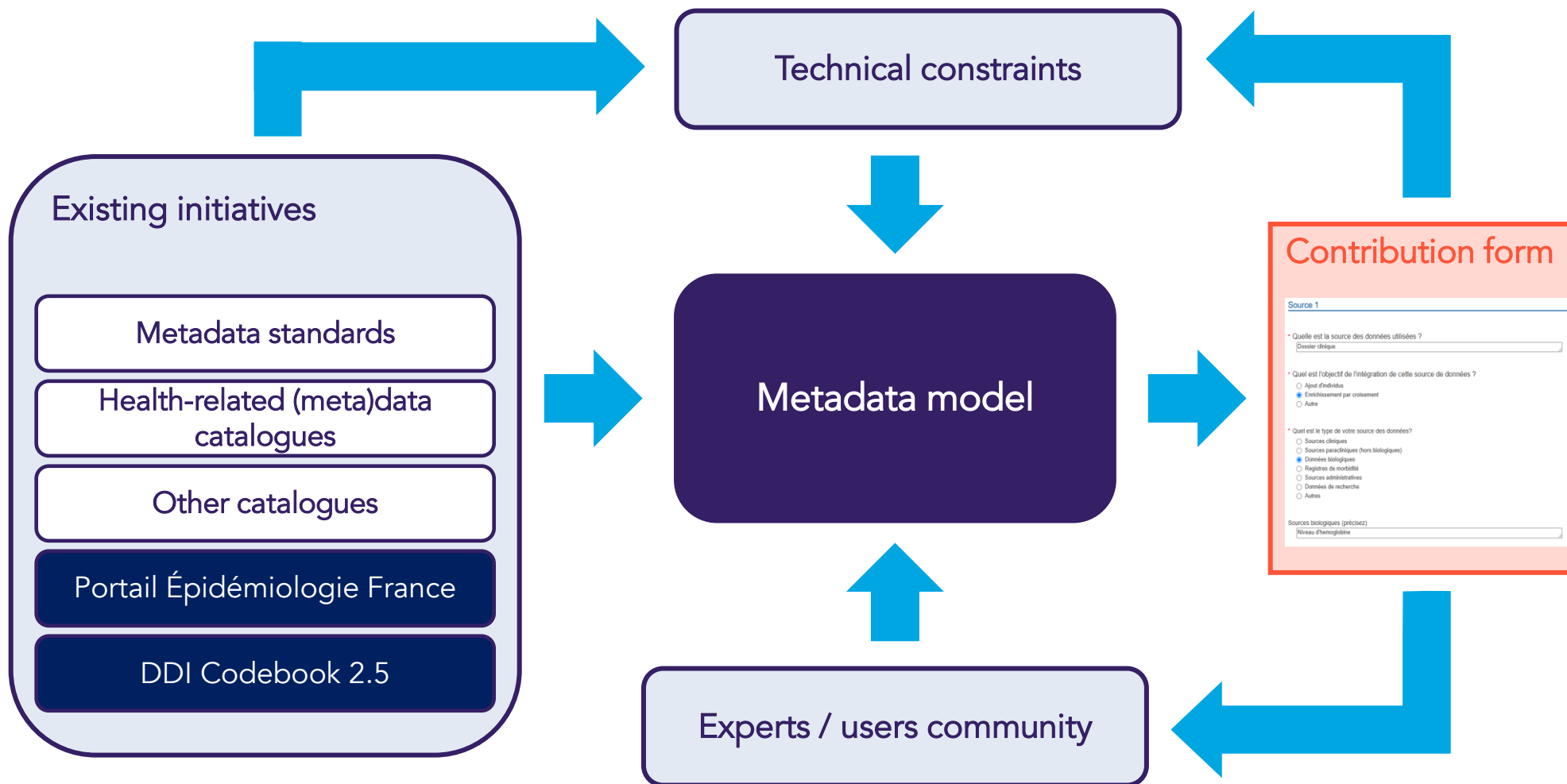
2031

Portal design and technical
development

Ensuring interoperability

Portal operation and management

The FReSH metadata model



How DDI helped us

Requirements

- Rely on **metadata standards** for interoperability
- Make use as much as possible of **open source tools**

DDI Codebook 2.5

- **Rich documentation**
 - Study context
 - Study methodology
 - Data access policies
 - Variables description
- **NADA Cataloguing tool**

The FReSH metadata model

Fields mapped on DDI-C

Study methodology	
Study type Interventional study Is this a clinical trial? Primary purpose Intervention model Allocation Type Unit Masking Type Details Arms/Groups Name Type Description Intervention Name Type Description	Analysis unit Trial phase Observational study Observational Study Design Time perspective Enrollement groups Name Description Exposure Studied factor, process, or exposure Name Type Description

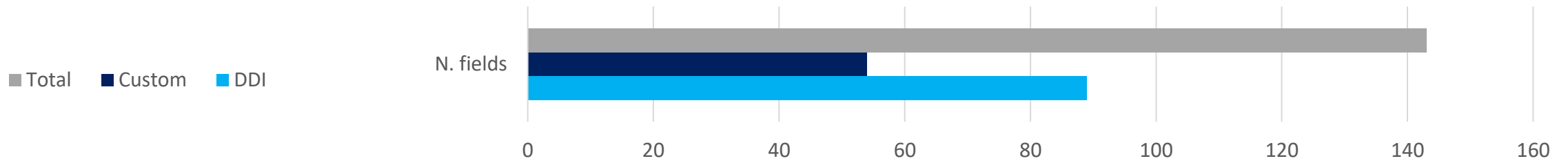
General description of the study	
Study overview Title Acronym Study Status Study purpose Summary Keywords	Study Population Population type Sex Age Inclusion criteria Exclusion criteria Countries covered Regions covered Geographical coverage details
Study Theme Medical Specialities Health determinants Conditions Rare diseases	

Data collection	
Primary data collection Collection Procedure Collection Mode Details Sampling Procedure at Inclusion Participant Recruitment Source Active Follow-up Data types	Data reused in the study Declaration of conformity Source Type Source Description Source ID Source Integration Purpose
Collection chronology Collection Start Date Collection End Date Collection Frequency	Number of participants Expected number Actual number

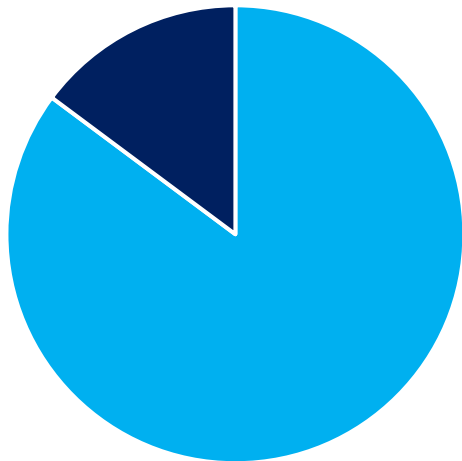
Administrative information		
PI Name Email Identifiers Affiliation organisation Name Identifiers Laboratory/Team Name Identifier Contact point	Team member Name Email Identifiers Affiliation organisation Name Identifiers Laboratory/Team Name Identifier Contact point	Sponsor Name Type Identifier
Funder Name Type Identifier	Regulatory requirements Authorisations/approval Other	Governance Sci or steering committee Details
Collaborations Networks, consortium Details	Study IDs ID Type Identifier	

Data access	
Data quality Standards or nomenclatures used Quality procedures used	Use conditions Data access conditions Access restrictions Link to sharing plan or additional information on data access Non-disclosure agreement Contact person for data information Name Email
Data availability Access to individual data Access to aggregated data Data access request tool Data file completeness Data location	Persistent identifier of the dataset
Obligations related to data use Reporting requirement Citation requirement	Data dictionary Mock sample
Additional documentation on data	

FReSH metadata model vs DDI-C: a few statistics

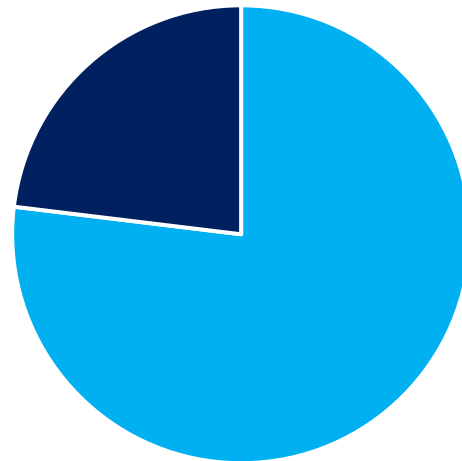


General description



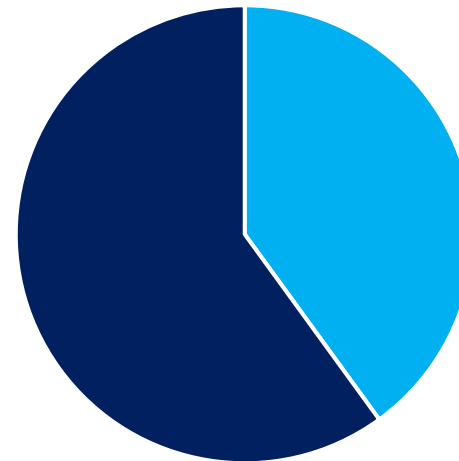
■ DDI ■ Custom

Administrative information



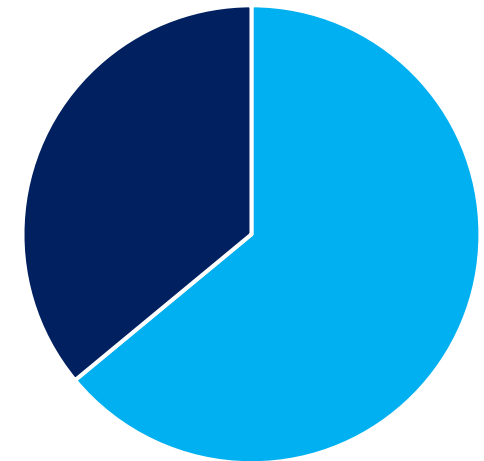
■ DDI ■ Custom

Study methodology



■ DDI ■ Custom

Data collection and access



■ DDI ■ Custom

Challenges encountered

Some DDI Codebook fields
were too broad for our
specific domain

- Using controlled vocabularies
(ex. ICD-11 pathologies)
- Labeling attributes

```
<topcClas  
  vocab="ICD-11"  
  vocabURI="http://id.who.int/icd/entity/11  
76751173">  
  Fear of disease  
</topcClas>  
  
<topcClas vocab="HealthDeterminants">  
  Environmental determinants  
</topcClas>
```

Challenges encountered

Some FReSH elements could not be mapped into DDI

- By using «notes» elements
- Some fields are not exported into DDI
 - JSON export

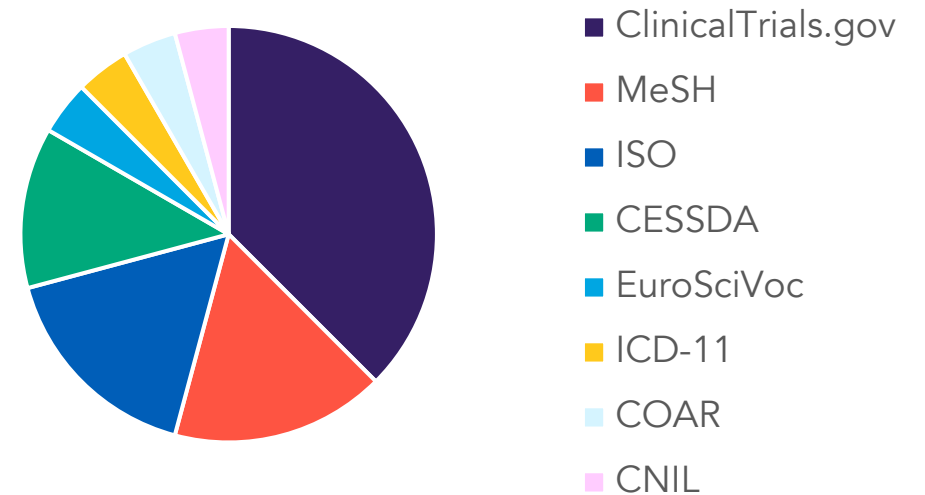
```
<!-- stdyDscr/method/dataColl/sources -->  
<notes subject="source purpose">  
    Data enrichment through cross-referencing  
</notes>  
...  
  
<!-- stdyDscr/method/ -->  
<notes subject="follow-up method">  
    Monitoring by participant visit to the  
    investigation center  
</notes>
```

Challenges encountered

Some existing terminologies
couldn't fit the needs of
FReSH

- Sets of terms of the existing CVs cover more than needed
 - Only a smaller set of terms is proposed to the user
- Term definition/label is too broad or unclear
 - Propose to the user a clearer definition/label
- Set of terms of the existing CVs lacked coverage
 - Added extra term to cover a domain-specific case

Source of existing vocabularies



Open questions

- How to define a «standard» DDI (Codebook) profile for the health domain?
- How to address interoperability with other domain-specific standards?
- How to address the problem of non-DDI fields?

Perspectives

- Documentation
 - <https://github.com/portail-fresh>
 - <https://github.com/portail-fresh/technical-documentation> (MkDocs)
- France Cohortes / FReSH: associate members
- Gather health-related actors around DDI (Codebook) community
- Definition of DDI profiles in French DDI (FrEDDI) WG
- Make proposals/suggestions for the CESSDA/DDI terminologies



Thank you!