Users' needs and research practices in data reuse: translating the FAIR principles in to reality

Roberta Ruggieri, Lucio Pisacane, Daniela Luzi

National Research Council, Institute for Research on Population and Social Policies, Rome, Italy

IRPPS
National Research Council
Institute for Research on Population and Social Policies

BACKGROUND

Within the OpenUP project a Pilot was performed in the field of Social sciences aiming to identify strengths and weaknesses in the process of dataset review and validation and outline best practices towards transparency, data dissemination, reliability and reuse.

To reach this aim we selected the scientific community that originated the Human Mortality Database (HMD), a well-known data source providing open access data for demography and population studies. We interviewed the HMD managers to investigate motivations and organizational features that prompted the community to make their data available as well as the quality assurance process. Moreover, to analyze attitudes and re-use behaviors a questionnaire was designed in collaboration with HMD management and sent to HMD end users.

The interviews highlighted some principles that guided the community manage and share data in a transparent way. They were: comparability, flexibility, accessibility and reproducibility. Survey's respondents provided a feedback on the database features they consider important as well as aspects that should be improved.

RESEARCH QUESTIONS

STEPS

Matching Providers' vision with User's perspective

To what extent users' requirement are fulfilled by the principles expressed by the HMD managers?

Do users' suggestions require the implementation of further principles or add features that complement the existing ones?

→ Principles that guided the development of HMD captured during interviews were matched with some responses on practices and appraisal given by HMD users in the survey

From the FAIR perspective
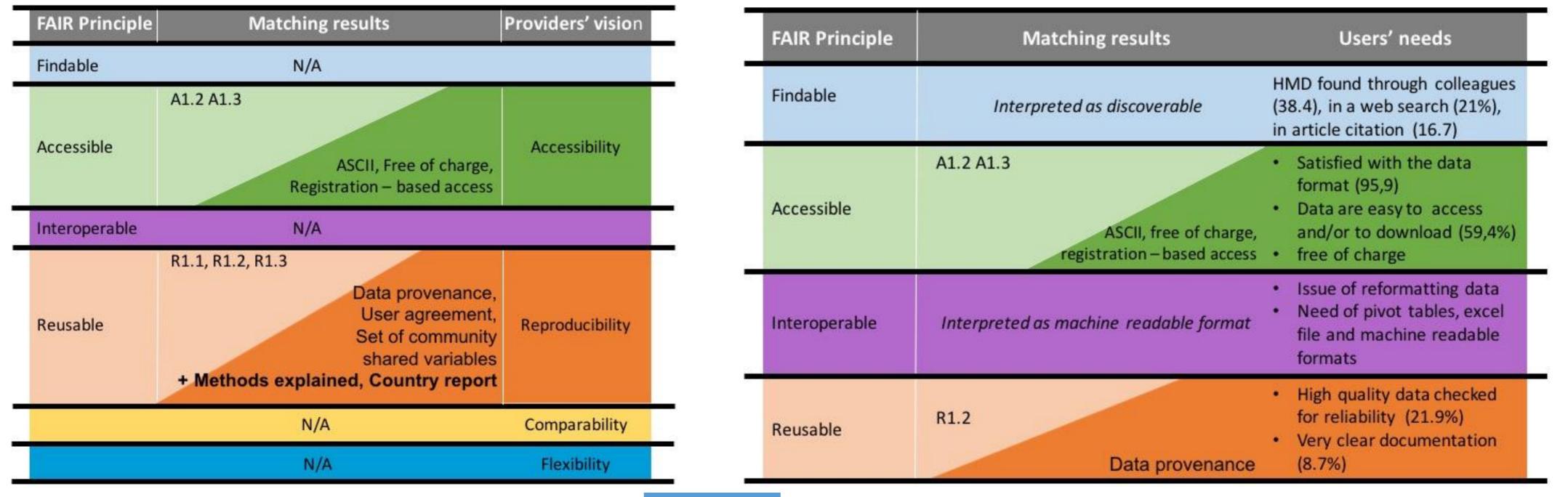
Do the Fair principles reflect the Provider's vision

Do the FAIR principles reflect User's needs?

→ Identification of the FAIR sub-principles that match the HMD providers' vision and users' needs.

Matching providers' vision with users' expectations

Provider's vision	User 's perspective	Additional features required by users
Accessibility • Free of charge access of data • Data available in a open, no-proprietary format (ASCII)	• "Free of charge access to data" • Data are easy to access and/or to download (59,4%) • Satisfied with the data format (95.9%)	• "Data in a text file need to be reformatted" • "My use of the data requires pivot tables" • "I prefer excel files"
Comparability • Uniform scientific methodology to format and process input data	• Comparability over time and across-countries (55.5%) • Long time periods of the data available (45.2%) • Access data related to all countries (45.6%)	
Flexibility • Uniform set of procedures tailored to the country' s history and socio-political development • Country report explains specificity and motivation of procedures applied	• Data are very detailed (20.5%) • Data are produced with reliable methods (15.0%)	Increase harmonization among documentation
Reproducibility • Availability of raw data • Explanation of method applied • Explanatory documentation • Country report	• The data are of high quality and have been checked for reliability (21.9%) • The documentation is very clear (8.7%)	• I prefer machine readable formats • Data automatic imports into statistical packages

Matching FAIR principles with the Providers' vision

FAIR Principle	Matching results	Providers' vision
Findable	N/A	
Accessible	A1.2 A1.3 ASCII, Free of charge, Registration – based access	Accessibility
Interoperable	N/A	
Reusable	R1.1, R1.2, R1.3 Data provenance, User agreement, Set of community shared variables **+ Methods explained, Country report**	Reproducibility
	N/A	Comparability
	N/A	Flexibility

Matching FAIR principles with Users' needs

FAIR Principle	Matching results	Users' needs
Findable	*Interpreted as discoverable*	HMD found through colleagues (38.4), in a web search (21%), in article citation (16.7)
Accessible	A1.2 A1.3 ASCII, free of charge, registration – based access	• Satisfied with the data format (95,9) • Data are easy to access and/or to download (59,4%) • free of charge
Interoperable	*Interpreted as machine readable format*	• Issue of reformatting data • Need of pivot tables, excel file and machine readable formats
Reusable	R1.2 Data provenance	• High quality data checked for reliability (21.9%) • Very clear documentation (8.7%)

RESULTS

• Comparability is the most appreciated feature by HMD users, reflecting one of the main principles that guided the HMD developers. Additional features that complement the HMD providers' vision are suggested mainly related to an automatic data transfer and elaboration.

• Accessibility and reproducibility are the common shared principles. HMD providers do not consider findability and interoperability as important issues. They identified two other principles comparability and flexibility, stressing importance of using a sound scientific method to provide mortality rates (i.e. Explanation of method and country reports).

• FAIR principles related to machine-centered aspects (i.e. persistent identifier, formal language for knowledge representation) do not seem to be a priority for HMD users (different interpretation). Some sub-principles of accessibility and reusability are mentioned in the users' survey, mainly related with data format and quality.

OpenUP HUB
https://www.openuphub.eu/
https://www.mortality.org/

Roberta Ruggieri: r.ruggieri@irpps.cnr.it