Datasharing in Nutrition: the pros and cons. How can we optimize research infrastructures to serve the needs of nutritional systems biology

Giuditta Perozzi[1], Irina Dobre[2], Raffaella Canali[1], Lars Dragsted[2] and the ENPADASI Consortium

[1]CREA-NUT - Food and Nutrition Research Centre, Roma, Italy; [2] University of Copenhagen, Denmark

Advanced computational analysis of shared datasets can increase single study power and provide a more complete overview of the health effects of dietary compounds, but awareness of this potential is still limited among nutritional scientists. The ENPADASI project of the Joint Programming Initiative "Healthy Diet for a Healthy Life" (www.enpadasi.eu) aims at establishing a nutrition-specific Research Infrastructure (RI), which builds on the Nutritional Phenotype Database (dbNP – www.dbnp.org), originally developed by NuGO and tested as a case study within the EU-FP7 EuroDISH project by combining metabolomics datasets from different intervention studies. The DASH-IN Infrastructure which is being developed in ENPADASI, connects datasets contributed by the partners from nutritional intervention/observational studies, but it is also open for data upload by others (www.enpadasi.eu/wp6/html).

Within the activities of WP2 we needed to analyze the perceived pros and cons of datasharing through the DASH-IN infrastructure, and to determine the expectations that will help its further development. For this purpose, an online survey was organized to guide the design of case studies to be performed during the project, and the link was distributed to all 51 ENPADASI partners spread over 9 EU countries. The questions contained in the survey revolved around 2 main questions:

1 – Why would you consider sharing data?
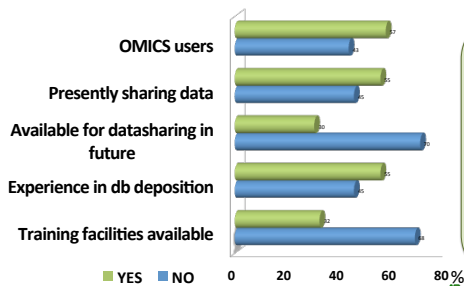2 – Why would you consider using shared data?

COUNTRY	# PARTNERS	RESPONDING PARTNERS N (%)	# SURVEYS
BE	13	6 (46%)	10
DE	7	7 (100%)	7
DK	1	1 (100%)	1
EE	2	2 (100%)	3
ES	3	2 (67%)	3
FR	2	1 (50%)	2
IE	4	3 (75%)	3
IT	18	14 (78%)	16
NL	1	1 (100%)	2
OVERALL	51	37 (73%)	47

The overall response rate was very high (73%), the non responders were partners with bioinformatic expertise, who are involved in developing the infrastructure but do not contribute datasets. Among the responding partners, slightly more than half employ high throughput –omics technologies, while the remaining half is composed of nutritional epidemiologist and public health nutritionist.
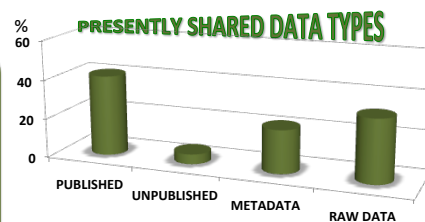
RESULTS QUESTION 1
Why would you consider sharing data?

DATASHARING EXPERIENCE

- OMICS users
- Presently sharing data
- Available for datasharing in future
- Experience in db deposition
- Training facilities available

YES ■ NO ■

TAKEHOME MESSAGE

- Over 50% of respondents already experienced deposition of their datasets in a research infrastructure, and correspondingly allowed datasharing
- Only 3 are willing to share unpublished data
- 14% of the partners who DO NOT presently deposit their data in dbs are willing to share them in the future.
- Db usage is also limited by lack of training facilities in 68% of the responders' Institutions

PRESENTLY SHARED DATA TYPES

PUBLISHED   UNPUBLISHED   METADATA   RAW DATA

Scientific journals appear as major drivers in datasharing: 52% of the respondents ahare their data to meet journal policies requesting data deposition prior to publication. At the same time, 28% do not share their data for fear of being scooped of the unpublished (28%). Howver, the perceived advantages from shared data analysis also provide strong motivations for sharing (33%).
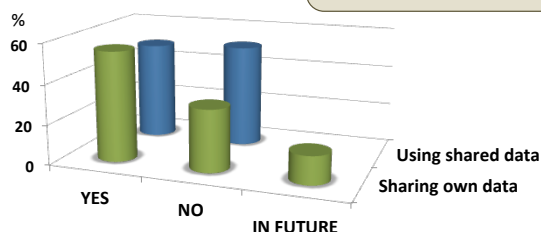
MAIN REASONS FOR SHARING/NOT SHARING

- Publication
- Data analysis
- Grant submission
- Other (unexplained)

SHARING ■ NOT SHARING ■

MAIN REASONS FOR SHARING/NOT SHARING

MOST FREQUENTLY USED dbs and RIs
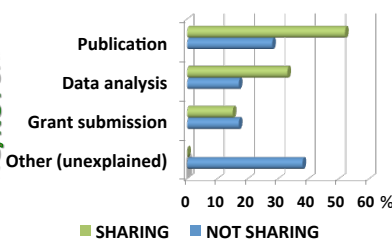
RESULTS QUESTION 2
Why would you consider using shared data?

Using shared data
Sharing own data

YES   NO   IN FUTURE

Almost 50% of ENPADASI partners have experienced use of shared data. The most frequent form of acknowledgement is co-Authorship (37%), especially for unpublished data, closely followed by citation of publications (34%) and Acession numbers (29) referring to the re-used datasets
The main reasons for NOT using shared data are based on issues of data quality, insufficient standardization, lack of specific expertise

CONCLUSIONS

Datasharing through nutrition-specific RIs can be strongly promoted by solving issues related to data quality, standardization, privacy and legal issues, as well as the need for training in the use of Ris, all aspects which are being addressed within the ENPADASI project.

These results point at crucial aspects that need to be solved to maximise the advantages of datasharing through nutrition-specific RIs, toward the long-term goal of achieving a fully open access environment supporting top level nutritional science.