



## OPINION ARTICLE

# The future of food and nutrition in ELIXIR [version 1; peer review: 2 approved with reservations]

Bachir Balech <sup>1</sup>, Lorraine Brennan<sup>2</sup>, Enrique Carrillo de Santa Pau <sup>3</sup>, Duccio Cavalieri<sup>4</sup>, Susan Coort <sup>5</sup>, Domenica D'Elia <sup>6</sup>, Lars Ove Dragsted <sup>7</sup>, Tome Eftimov<sup>8</sup>, Chris T. Evelo <sup>5</sup>, Polonca Ferk<sup>9</sup>, Paul Finglas<sup>10</sup>, Agnese Gori<sup>4</sup>, John Hancock<sup>11</sup>, Matúš Kalaš <sup>12</sup>, Barbara Koroušić Seljak<sup>8</sup>, Carl Lachat <sup>13</sup>, Brane Leskošek<sup>9</sup>, Edoardo Pasolli<sup>14</sup>, Graziano Pesole <sup>1,15</sup>, Karl Presser<sup>16</sup>, Anna Sandionigi <sup>17,18</sup>, Monica Santamaria <sup>1</sup>, Duygu Dede Şener<sup>5</sup>, Maria Traka <sup>10</sup>, Guy Vergères<sup>19</sup>, Karin L. Zimmermann<sup>20</sup>, Jildau Bouwman <sup>21</sup>

<sup>1</sup>Bioenergetics and Molecular Biotechnologies, Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, Bari, 70126, Italy

<sup>2</sup>UCD Institute of Food and Health, UCD School of Agriculture and Food Science, Dublin, 4, Ireland

<sup>3</sup>Computational Biology Group, Precision Nutrition and Cancer Research Program, IMDEA Food Institute, Madrid, Spain

<sup>4</sup>Department of Biology, University of Florence, Firenze, 50019, Italy

<sup>5</sup>Department of Bioinformatics - BIGCaT, NUTRIM Research School, Maastricht University, Maastricht, The Netherlands

<sup>6</sup>Consiglio Nazionale delle Ricerche, Institute for Biomedical Technologies, Bari, 70126, Italy

<sup>7</sup>Department of Nutrition, Exercise and Sports, University of Copenhagen, Copenhagen, Denmark

<sup>8</sup>Computer Systems Department, Jožef Stefan Institute, Ljubljana, SI-1000, Slovenia

<sup>9</sup>Faculty of Medicine, Institute for Biostatistics and Medical Informatics, University of Ljubljana, Ljubljana, Slovenia

<sup>10</sup>Food Databanks National Capability, Quadram Institute Bioscience, Norwich, NR4 7UQ, UK

<sup>11</sup>Faculty of Medicine, Institute of Biochemistry, University of Ljubljana, Ljubljana, Slovenia

<sup>12</sup>Computational Biology Unit, Department of Informatics, University of Bergen, Bergen, Norway

<sup>13</sup>Department of Food Technology, Safety and Health, Ghent University, Ghent, Belgium

<sup>14</sup>Department of Agricultural Sciences,, University of Naples Federico II, Portici, 80055 Portici, Italy

<sup>15</sup>Department of Biosciences, Biotechnology and Biopharmaceutics, University of Bari, Bari, Italy

<sup>16</sup>Premotec GmbH, Winterthur, Switzerland

<sup>17</sup>Department of Informatics, Systems and Communication, University of Milan-Bicocca, Milan, Italy

<sup>18</sup>Quantia Consulting srl, Milan, Italy

<sup>19</sup>Food Microbial Systems Research Division, Agroscope, Berne, CH-3003, Switzerland

<sup>20</sup>Research group Consumer and Chains, Wageningen Economic Research (WEcR), The Hague, 522595 BM, The Netherlands

<sup>21</sup>Microbiology and Systems Biology, TNO, Zeist, 3704HE, The Netherlands

**V1** First published: 25 Aug 2022, 11(ELIXIR):978  
<https://doi.org/10.12688/f1000research.51747.1>

Latest published: 25 Aug 2022, 11(ELIXIR):978  
<https://doi.org/10.12688/f1000research.51747.1>

## Abstract

Non-communicable diseases are on the rise and are often related to food choices; nutrition affects infectious diseases too. Therefore, there is growing interest in research on public and personal health, as related to food, nutrition behaviour and well-being of consumers throughout the life cycle. These concepts and their relations are

## Open Peer Review

Approval Status ? ?

	1	2
version 1 25 Aug 2022	? view	? view

complex and only partially understood – more data is needed to improve our understanding. The required data include deep genotyping and phenotyping data from human nutritional studies, covering metabolic and health, but also including behavioural and socio-economic data. The aim of this white paper is to provide a guiding framework for the future contributions of ELIXIR's recently established Food & Nutrition (F&N) Community. This white paper is the direct result of a strategy meeting that took place in September 2019 in The Hague (NL) and involved representatives of 14 countries representing the ELIXIR Nodes. The meeting led to the definition of F&N related bioinformatics challenges, including the use of standards for data reuse and sharing, and for interoperability of data, tools and services, advocacy and training. Resolving these bioinformatics challenges makes it possible to address a wide range of F&N-related challenges, such as definition of an individual health status, individual dietary needs, and finding complex intake biomarkers (to replace questionnaires). Here, we provide suggestions on how to align these tasks within the ELIXIR Platforms, other ELIXIR Communities/Focus Groups and the European Strategy Forum on Research Infrastructures.

#### Keywords


Food, nutrition, ELIXIR, data reuse, bioinformatics, interoperability




This article is included in the [ELIXIR](#) gateway.



This article is included in the [Agriculture, Food and Nutrition](#) gateway.

1. **Graham King** , Southern Cross University, Lismore, Australia

2. **Anu Bansal** , Lovely Professional University, Phagwara, Punjab, India

Any reports and responses or comments on the article can be found at the end of the article.

**Corresponding author:** Jildau Bouwman ([jildau.bouwman@tno.nl](mailto:jildau.bouwman@tno.nl))

**Author roles:** **Balech B:** Conceptualization, Methodology, Writing – Review & Editing; **Brennan L:** Conceptualization, Methodology, Writing – Review & Editing; **Carrillo de Santa Pau E:** Conceptualization, Methodology, Writing – Review & Editing; **Cavalieri D:** Conceptualization, Methodology, Writing – Review & Editing; **Coort S:** Conceptualization, Methodology, Writing – Review & Editing; **D’Elia D:** Conceptualization, Methodology, Writing – Review & Editing; **Dragsted LO:** Conceptualization, Methodology, Writing – Review & Editing; **Eftimov T:** Conceptualization, Methodology, Writing – Review & Editing; **Evelo CT:** Conceptualization, Methodology, Writing – Review & Editing; **Ferk P:** Conceptualization, Methodology, Writing – Review & Editing; **Finglas P:** Conceptualization, Methodology, Writing – Review & Editing; **Gori A:** Conceptualization, Methodology, Writing – Review & Editing; **Hancock J:** Conceptualization, Methodology, Writing – Review & Editing; **Kalaš M:** Conceptualization, Methodology, Writing – Review & Editing; **Koroušič Seljak B:** Conceptualization, Methodology, Writing – Review & Editing; **Lachat C:** Conceptualization, Methodology, Writing – Review & Editing; **Leskošek B:** Conceptualization, Methodology, Writing – Review & Editing; **Pasolli E:** Conceptualization, Methodology, Writing – Review & Editing; **Pesole G:** Conceptualization, Methodology, Writing – Review & Editing; **Presser K:** Conceptualization, Methodology, Writing – Review & Editing; **Sandionigi A:** Conceptualization, Methodology, Writing – Review & Editing; **Santamaria M:** Conceptualization, Methodology, Writing – Review & Editing; **Şener DD:** Conceptualization, Methodology, Writing – Review & Editing; **Traka M:** Conceptualization, Methodology, Writing – Review & Editing; **Vergères G:** Conceptualization, Methodology, Writing – Review & Editing; **Zimmermann KL:** Conceptualization, Methodology, Writing – Review & Editing; **Bouwman J:** Conceptualization, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** LOD has research contracts with foundations owning larger food or pharma companies, currently the Carlsberg Foundation and the Novo Nordic foundation. JB is working for TNO, a non-for-profit research institute that also does research for companies. Other authors declare no conflict of interest.

**Grant information:** The author(s) declared that no grants were involved in supporting this work.

**Copyright:** © 2022 Balech B *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Balech B, Brennan L, Carrillo de Santa Pau E *et al.* **The future of food and nutrition in ELIXIR [version 1; peer review: 2 approved with reservations]** F1000Research 2022, **11**(ELIXIR):978 <https://doi.org/10.12688/f1000research.51747.1>

**First published:** 25 Aug 2022, **11**(ELIXIR):978 <https://doi.org/10.12688/f1000research.51747.1>

## Introduction

Europe is faced with a range of food-, nutrition- and health-related challenges, often embedded in socioeconomic inequalities.<sup>1</sup> Obesity levels and chronic diseases such as type II diabetes, cardiovascular diseases, dyslipidaemia, increased allergies, asthma, and neurodegenerative diseases, and certain forms of cancer, are often related to those food choices. Moreover, healthy, and sustainable diets are a cornerstone included in the “From Farm to Folk” European Union strategy for the European Green Deal to reduce the risk of life-threatening diseases and the environmental impact of our food system. Therefore, there is growing interest in research on public and personal health, as related to food, nutrition behaviour and well-being of consumers throughout the life cycle.

Resolving the above issues require understanding of food choices, food intake, food composition and the effect of nutrition on health. These concepts and their relations are complex and only partially understood so more data is needed to improve our understanding. The required data include deep geno- and phenotyping data, from human nutritional studies covering metabolic and health, but also including behavioural and socio-economic data. These big (and many small) data represent an opportunity for the development of products and interventions that help shift current consumption patterns towards more healthy and sustainable diets for each citizen.

The Food and Nutrition (F&N) Community is well organized and has already collated datasets, FAIRified several of its data sources and developed tools (see [Table 3](#)) that would be immediately available for the meaningful harmonization, integration, and analysis of these datasets. Several efforts have collected and are collecting datasets in a FAIR way ([ELIXIR Position Paper](#)) and many semantic resources, such as ontologies, have been built, for instance in the EU-funded projects [ENPADASI](#), [Richfields](#), [EuroDISH](#), [JPI HDHL INTIMIC](#) knowledge platform, and the [FNS-Cloud](#) (see [Table 2](#)). Furthermore, a dedicated community is active and organized in [NuGO](#), an association of universities and research institutes focusing on the joint development of the research areas of molecular nutrition, personalised nutrition, nutrigenetics, nutrigenomics, nutriepigenomics and nutritional systems biology. A part of the community is also organised in [EuroFIR AISBL](#) an international non-profit association ensuring sustained advocacy for food information in Europe. The community is working towards an ESFRI research infrastructure: [Food, Nutrition and Health RI](#). However, to date, efforts are still fragmented, and the integration requires a wide range of data science [1], better ontologies, improved databases, data harmonization, algorithms and tools for easy sharing and dissemination of the data, while respecting data protection policy. Collaborations between dieticians, nutritionists, bioinformaticians, biostatisticians, systems biologists, consumer scientists, data scientists and knowledge engineers are needed for an interdisciplinary approach to nutrition issues.

To address these challenges, we organized a workshop “ELIXIR Food and Nutrition Community Workshop” with independent food, nutrition, bioinformatics, systems biology, data and computer science experts, and those within [ELIXIR](#). This opinion article summarizes the interactions in the workshop and its outcomes and describes the potential role of the F&N Community in relation to [ELIXIR](#).

[ELIXIR](#) coordinates bioinformatics resources across its member states and helps researchers to find, analyse, and exchange biological data. It is a distributed infrastructure with a Hub based in Hinxton, United Kingdom, and an increasing number of Nodes located throughout Europe. As of June 2020, [ELIXIR](#) has 22 national Nodes, with European Bioinformatics Institute (EMBL-EBI; co-located with the Hub), working as a separate Node.

### Workshop “ELIXIR Food and Nutrition Community Workshop”

The workshop was organised on the 23<sup>rd</sup> and 24<sup>th</sup> of September 2019 in The Hague (The Netherlands). The invitation to the workshop was widely advertised through the newsletter of the [NuGO](#), the [ENPADASI](#), [Richfields](#) and [FNH-RI](#) mailing lists, and [ELIXIR](#) dissemination channels, including [ELIXIR](#) Technical Coordinators, Heads of Node mailing lists and the [ELIXIR newsletter](#).

The workshop included 26 participants from across 14 countries (NL, DE, BE, SI, UK, SE, DK, ES, CH, IT, FR, FI, EE, IE) representing the [ELIXIR](#) Nodes, including the [ELIXIR](#) Hub, and additional countries. The objective of the meeting was to identify the principal challenges in food and nutrition and prioritise actions, in particular those within the scope and mission of [ELIXIR](#). The workshop showcased flash presentations on the (inter) national F&N activities, ideas, and requirements that an [ELIXIR](#) F&N Community could address. An overview of [ELIXIR](#) Use Cases and Platforms was also presented by a representative of the [ELIXIR](#) Hub. The [ELIXIR](#) Training Platform was presented by one of the Training

<sup>1</sup>Data Science can have different specialisations, depending on the domain of research; it is used in fields like bioinformatics and biostatistics, and in other areas as well like computer science, biology, biochemistry, medicine, statistics, math and engineering. Data Science uses the tools, code them or even develop new and better data models and algorithms.

Platform coordinators. The presentations were followed by discussions on the needs and challenges present in the F&N community. The following challenges were identified:

- In order to measure a health effect, the **individual health status** needs to be defined
- It is clear that every **individual** has different **dietary needs**, but what advice is for an individual is still hard to define (e.g. what advice is needed for an individual at risk for a non-communicable disease)
- Questionnaires are now the standard to quantify what individuals have eaten. These measures are very unreliable. **Complex dietary and food intake biomarkers** may make these measures more reliable.

These F&N challenges can only be efficiently resolved if a number of bioinformatic challenges are overcome:

- Standardization/interoperability
  - Terminology: ontology development (not all relevant ontologies for F&N are in place) to make data interoperable
  - Standardization of questionnaires (questionnaires are rarely aligned)
  - Tools for standardization to result in homogeneous data and metadata (tools for standardization are not always available to F&N scientists)
  - Technical interoperability solutions
  - Training/Capacity Building data stewardship, data annotation/curation, etc.

#### Data availability

- Findability, Accessibility and Connectivity of datasets (Diet intake, Health, Food composition, Consumer data, and Omics data): all shared data should be (machine) Findable and Accessible
- Rich meta-data capture (not all datasets have complete meta-data to be Reusable)
- Ethical compliance and approval in case of person-identifiable data (including consent: authorisation and authentication including permissions registry and confirmation of informed consent and added purpose binding)
- Data reuse
  - Data integration (including omics and personalization) with a specific focus on metabolomics and microbiomics (but proteomics, epigenetics and transcriptomics are relevant to the community as well)
- Advocacy and Training
  - Researchers and governmental organisations publish research documents to spread knowledge but hardly publish data (raw or modified). Convincing arguments and incentives must be created to consider publication of documents of equal importance as publication of data.
  - Training/Capacity Building on FAIR data and standards (not all F&N scientists know the FAIR principles and are aware of the standard's importance to align their data management plans with FAIR principles and where to find support)

- Tool and service availability and interoperability
  - Make tools and services available with which data can be analysed, visualised and manipulated
  - Need for alignment of existing and new analysis software with ELIXIR existing ELIXIR efforts (e.g. microbiomics/[GALAXY](#) and similar dataflows)
  - Public and private repositories must be integrated in such a way that it allows users to easily transfer data into existing tools for their data processing. This would lead to a landscape of repositories and tools where an arbitrary number of systems can be connected or chained to perform data analysis.
- Networking actions
  - Interaction with consumers (onboarding and transparency) and other stakeholders such as F&N researchers, policy makers, educators, industry, hospital and patients
  - Alignment with European Open Science Cloud (EOSC) strategy for sustainable long term data reuse and other initiatives.

These challenges are described in more detail in the following sections.

## Food and nutrition challenges

### *Individual health concept*

Being healthy as opposed to having a disease, is generally not the way healthiness is defined today. It has been accepted by the [World Health Organization](#) that being healthy also includes physical, mental and social well-being. This indicates the need for new ways to quantify health status and include relevant data on well-being and in healthy people. How can we go from general health advice to subgroup or even to personal advice, linking with a human data use case? Can we use dynamic data for this? Can we set up methods to measure health status in relation to nutrition?

In order to measure health we need to operationalize the concept of health and to make it measurable. In the F&N community this is often done by challenge tests to probe the individual resilience to a specific health challenge. The idea behind challenge tests is that a dynamic response is telling more about the status of the underlying processes than only the baseline values. Challenge tests are therefore a way forward in quantifying individual health.

Ideally, we would like to know all relevant aspects of health or at least the most important ones for general metabolic balance but this may seem like a difficult target to reach. Using omics, it is potentially possible to measure multiple responses to a complex health challenge and to span the individual responses across many metabolic pathways covering the human as well as the microbial composition and metabolism. This approach would circumvent the need for pre-defining all aspects of health while allowing actual variability in health to be determined. Exposing individuals to a multitude of different nutritional challenges is therefore a prerequisite for measuring nutritional health. A diet is in fact such a multifaceted health challenge and extracting information on health status is therefore likely embedded in many multi-omics nutritional datasets. For example, the variation of gut microbiome composition, often in response to external cues such as diet,<sup>2</sup> and its far-reaching effects on the host in highly prevalent disorders is widely documented in the literature.<sup>3-6</sup> In addition, the microbiome is emerging as an important determinant of metabolic responses to food and a driver for inter-individual variability in metabolic health biomarkers.<sup>7,8</sup> For a more meaningful interpretation of the mechanisms underlying this two-way interplay and for translating microbiome research into effective benefits for the host, a correlation with other -omics profiles, such as metabolome<sup>9</sup> or host transcriptomes,<sup>10,11</sup> through physiological and pathological conditions, is required. Relatively few such datasets have been shared and open access to varied nutritional data is therefore an important goal. For this purpose, both data on diseased and healthy citizens are needed. This case study is focused on connecting the human nutritional datasets available in the F&N community to the datasets available in ELIXIR (including the BBMRI and ECRIN data) that are more focused on diseased subjects.

Nutrigenomic biomarkers are needed to clarify association between the personal genetic background, phenotype, food, and nutrition. Additionally, interactions between drugs/pharmacotherapy/toxicology and food/nutrition should be investigated.<sup>12,13</sup>

To accomplish this, datasets from complex nutritional challenge tests and dietary interventions covering multiple foods, nutrients and common food additives and toxicants are needed and need to be integrated in smart ways with the current knowledge. Work in this area has been initiated years ago in collaborative projects under NuGO, however a generalized model for data interpretation has not yet been developed. Currently additional steps are being made in the projects FNS-cloud and JPI HDHL INTIMIC knowledge platform (JPI KP).

A considerable number of highly controlled meal studies and dietary intervention studies exist (e.g. see [Phenotype database](#) and trials registries) and data from these studies are in principle decided by the consortia to be FAIR; however, they are not yet fully FAIRified and efforts to provide such datasets including related rich omics data are therefore a major priority. However, rich metadata for data intensive omics, such as the microbiome, are an issue, in particular as comprehensive (rich) diet-related metadata are challenging to obtain routinely in nutritional intervention studies. In the JPI KP rich meta-data on the nutritional studies are integrated with metabolomics and microbiome data. This type of data (metabolomics and microbiome) are also available within ELIXIR and solutions are available to share them. Integrating the [Phenotype database](#) with these solutions, e.g. Metabolights would improve the current solutions and would make the nutritional data available to ELIXIR.

In addition, omics analysis software (and hardware) for optimal data extraction, standardization, integration, and analytics tools to address the individualized response across multiple pathways are needed for the F&N community. Several of these tools exist, but some are fragmented or in an immature form and need to be joined to comprehensive workflows, where ELIXIR may help.

### *Individual dietary needs*

Beyond defining individual health and resilience there is a need to develop methods to deliver dietary advice at an individual level to improve health. Data integration from a multitude of sources and different data types is crucial in understanding the individual human phenotype and what is optimal for his/her health. This challenge requires integration of data on behaviour, dietary intake, health-related endpoints, biometric dynamics, images, and omics data. Combining this information from a rich data source could provide a prediction of the longer-term health effect of a specific food or dietary intake even at the level of an individual.

Understanding and measuring food intake and its dynamics is central for developing effective lifestyle interventions. It is also relevant for the food industry, which has agreed to work with EU member states to make food supply healthier through food reformulation. Several tools and eHealth solutions relevant to a broader community and food industry have been developed. For instance, RICHFIELDS, JPI HDHL Foodball, ENPADASI, EIT Food, and FNS-Cloud have delivered the standardized requirements and ontologies<sup>14</sup> that were needed to unambiguously describe the determinants of consumers' food choices, the [composition of food](#) and dietary intake (including biomarkers of intake) as well as the subsequent metabolic effects of food components in the human body. A diversity of tools is needed to integrate (personal) data that can be linked to dietary, health, and behavioural (consumer) reference data and ontologies. To give personal advice, a combination of user-friendly apps with sensors to capture non-invasive or minimally invasive biometric signals in real time are needed, that are able to onboard individuals and keep them connected and motivated. Individual data will be generated that can be used to give personalized advice but can also be used from research. For secondary use of data, consent is needed, unless data are completely anonymized. To adhere to privacy legislation federated learning solutions may be needed. To support the reuse of these individual data, the infrastructure to support this is being developed where personal health and consumer study data will be stored safely (FNH-RI). ELIXIR has the knowledge and tooling that can help to develop the solution mentioned.

### *Complex intake biomarker*

All the above challenges require data on what people have eaten and their responses. Many self-reported questionnaires/diaries exist to enable collection of these data and some solutions are developed to make this data interoperable. However, it is important to recognise the limitations associated with such self-reported data: participants tend to under-report their food intake and the approaches are burdensome on participants thus resulting in issues with incomplete data.<sup>15,16</sup> Therefore, there is a great need for molecular biomarkers for food intake (BFIs) to provide stronger tools for intake assessment and understand the relationship to biomarkers of health and disease. A considerable number of such biomarkers have been identified<sup>17,18</sup> and in the JPI FoodBall project a systematic approach was used to define,<sup>19</sup> search<sup>20</sup> and validate<sup>21</sup> BFIs and to introduce elements of an ontology for the area.<sup>22</sup> Moreover, several food challenge studies were conducted to find new candidate BFIs for the most common foods in Europe, covering specific meats,<sup>23,24</sup> fruits,<sup>25,26</sup> vegetables,<sup>27</sup> legumes,<sup>28</sup> and several others. However, biomarkers for many foods have still not been proposed

and many of the candidate biomarkers published are still not fully validated. An example of an important aspect of validation, which is often missing in the work on BFIs is the quantity of the food ingested; the current methodology is largely qualitative and combining data across several studies could be a facile way to improve this aspect of validation. To develop and validate this type of biomarker rich standardized questionnaires on intake, as the (not-so-) gold standard, and metabolomics data (as potential source of the biomarkers) are often used,<sup>29</sup> but as more and more markers are identified they may also be further refined by biomarker combinations.<sup>26,30</sup> Work on biomarker approaches to assess whole diets have also been advanced in recent projects and could develop into another set of tools in the intake biomarker toolbox.<sup>31–33</sup> However, tools to search information related to biomarker validation and rich databases of nutritional studies with associated metabolomics data are needed to reach a coverage of the diet with appropriate biomarkers to allow improved assessment of dietary intakes and compliance. These data need also to be made interoperable, both in terms of the metadata provided with studies and the information needed to identify the compounds identified as potential BFIs. This requires rich databases of mass spectral and other compound information that can be searched and used for verification. Some of this work has been started by other international players and some of it in the JPI HDHL Foodball project, but the solutions and resources require further solutions for interoperability. The interaction with ELIXIR may make further development possible especially by connecting to the MetaboLights initiative or similar repositories.

## Bioinformatics challenges

### Standardization/interoperability

A key issue is to understand the relationships between consumer behaviour, food intake, and nutritional status to help consumers to choose a healthy diet. This information would also be helpful to commercial companies in the development of more healthy foods and dietary solutions. These types of data are collected by dietary monitoring systems in interventions and observational studies conducted across Europe using a standardized methodology. Important in this data collection is storage of relevant metadata on the study design as well as the methods used. Moreover, standards for dietary assessment, measurements for dietary intake and nutritional status are needed for comparability and quality of data, e.g. to start the discussion about FAIR (Findable, Accessible Interoperable and Reusable) food data. Guidelines to improve reporting of dietary assessment and nutrition research will help to re-use existing evidence for public health recommendations.<sup>34</sup> Relevant ontologies for nutrition knowledge can guide users towards better use and appraisal of research findings.<sup>35</sup>

In the last decades, a great amount of work has been done in predictive healthcare using Artificial Intelligence methods as a result of the existence of publicly available biomedical vocabularies and standards together with tools for standardization of health-related data. Despite the large number of resources in the health domain, the food and nutrition domain is still low-resourced. There exist only a few food ontologies that are developed for a specific application scenario, with a small number of studies that are focused on exploring relations between different ontologies and standards. For this reason, the workshop “**Big Food and Nutrition Data Management and Analysis - BFNDMA**” was initiated at the 2019 IEEE International Conference on Big Data (Los Angeles, USA) and is an ongoing initiative, with a focus on methodologies for management and analysis of food and nutrition data.

The project RICHFIELDS started working on standardization of food items that are described and classified using different standards, by presenting a method known as StandFood<sup>36</sup> that is a synergy of Natural Language Processing and Machine Learning in order to standardize and classify foods with regard to the FoodEx2 provided by the European Food Safety Agency (EFSA). The method focuses only on identifying lexical similarity between the English names of the food items. Further, this work continues as a part of the current project FNS-Cloud, where different food semantic standards (i.e. Hansard taxonomy, Food-On, OntoFood, and SNOMED-CT) have been explored for food data annotations. The results showed that not all ontologies relevant for F&N are in place and that the current ontologies do not have a good coverage.<sup>37</sup> Based on the results, the FoodOntoMap,<sup>38</sup> a semantic resource was created, which provides links between different food ontologies that can be further reused to develop applications for understanding the relation between food systems, human health, and the environment. Additionally, the FoodBase<sup>39</sup> corpus is one of the first annotated recipe corpora with food entities standardized using the Hansard taxonomy. To make this information available for subject-matter experts, the FoodViz<sup>40</sup> tool has been presented as a visualization tool for presenting automatically annotated food-related textual data, where subject-matter experts can check the automatically annotated results and also make corrections (i.e. manual annotations). In this way, textual data related to food, nutrition, and health, as a typical example of unstructured big data being collected from different data sources, can be handled.

Together with ELIXIR, the F&N Community will standardize, connect, and model the data pipeline. Furthermore, standardisation in food behaviour data (why do I eat what I eat), especially food choice motives, will provide tooling to create unique data pipelines based on linkage between food intake and its determinants including contextual variability and will prevent high measurement errors and time-consuming data collection. All tools and services will also become available for the wider ELIXIR community.



One of the aims of the interaction between ELIXIR and the F&N Community may be to unify the RICHFIELDS, ENPADASI and FNS-Cloud requirements and ontologies and link them with other standardised ones (provided through ELIXIR). Standardisation in consumer and human nutrition science is mandatory to exchange data and to monitor and analyse food behaviour. Data gathered by the ELIXIR Implementation Study “[A microbial metabolism resource for Systems Biology](#)”, are of special interest in this respect, as they can be exploited to integrate metagenomics data with the human nutritional phenotype. The effect of food in modulating the human gut microbiome is broadly recognized, but standardized tools for exploring big data at the interface between nutrition and the microbiome for self-sustainability of health are still fragmented and lack acceptable levels of standardization and integration.

The F&N Community is now using a number of profiling tools to explore the relationship of nutritional change with biological outcomes. This clearly aligns the area with a number of topics covered in ELIXIR. This includes data processing and analysis tools, databases for proteomics, genomics, mass spectral information and microbiomics. Especially for development of biomarkers related to health, dietary exposures and biological effects there is a good prospect for cross-fertilization of the food and nutrition area with several other topical areas within ELIXIR.

### Data availability

Some of the datasets in the F&N Community are already well structured, but not yet fully FAIR. For example, ENPADASI has developed three templates for data collection covering all types of study design, study content and study objectives. The project has collected 111 intervention studies (of which 27 studies were made open access) and 23 observational studies. Those studies include clinical chemistry data, transcriptomics, genomics, metabolomics and microbiome data. As databases the Phenotype database and the Opal/Mica systems were used. Datashield was used to integrate data from the different systems and the FAIR principles were considered as much as possible. Minimal requirements for nutritional data sharing and information on studies needed for quality appraisal were developed, these requirements will make sure that all information relevant to judge the study quality are shared, where possible. In addition, based on the templates and uploaded studies nutritional terms were identified that were mapped to existing ontologies and new ontologies are being developed for nutritional terms ([ONS](#)) and nutrition epidemiology ([Ontology for Nutritional Epidemiology](#)). As indicated, we need to connect to other data sources and knowledge to fulfil the goals in the community. This requires full implementation of FAIR and Omics related standards in ELIXIR.

RICHFIELDS worked towards solutions to collect data on dietary habits. This includes privacy sensitive data. Therefore, RICHFIELDS has examined how the GDPR Regulation addresses these privacy matters. RICHFIELDS provides a framework for the design of the ethical and legal aspects and includes the following recommendations: (1) use of pseudonymisation with appropriate safeguards for unauthorised reversal of pseudonymisation; (2) use of appropriate technical and organisational measures to ensure GDPR compliance; (3) systems for dealing with queries and requests from data subjects; (4) appointment of a Data Protection Officer (DPO); (5) mechanisms for handling freedom of information (FOI) requests; (6) use of suitable data protection clauses for trans-border data transfer; (7) obtaining insurance to cover liability in the event of data breaches; and (8) the establishment of an independent ethics committee with the remit to monitor the activities, protocols on matters relating to security, transfer of data to third countries, assessing genuineness of requests from data users and procedures for dealing with ethically suspect requests, and procedures for handling requests from data subjects. This framework may be integrated with the solutions on this matter in ELIXIR.

### Data reuse

The above-described F&N challenges all require complex interaction of diverse data sets. This integration will require knowledge networks, specific tools and algorithms development and analysis pipelines and may require hardware (cloud solutions). Although some Food and Nutrition knowledge networks and analysis pipelines are in place (e.g. [Micronutrients Wikipathways](#), [NutriGenomeDB](#)), many are still lacking or require better generalization for reuse. Currently the F&N community has no cloud solution available, which may be needed especially if microbiome data (or other datasets that require large computation volume) are being integrated. Several of the datasets needed in the community will be private data, for which explicit consent is needed if the data is reused. Solutions for requesting secondary data reuse are therefore important and need to be outlined to improve ease of application.

### F&N community in ELIXIR

Understanding the relation between diet, microbiome, metabolome and health was identified by popular vote as the one area where:

In our first implementation study we will first work towards a list of available datasets and tools within our community and then connect the currently available F&N databases to Metabolights.

## ELIXIR technical activities

ELIXIR technical activities are performed by ELIXIR Nodes and supported by the Hub. The Nodes run bioinformatics resources and services focusing on national priorities and contributing to a harmonised strategy across Europe. At the national level, an ELIXIR Node consists of research institutes within a member country, building on their national strengths. At the European level, ELIXIR's activities are structured around Platforms and Communities, which bring resources and expertise together from several ELIXIR Nodes. The ELIXIR Platforms are responsible for the implementation of the [ELIXIR Scientific Programme](#), which is organised into five key areas: Data, Tools, Compute, Interoperability and Training.

The 11 Communities that are currently recognised are: Federated Human Data, Rare Diseases, Human Copy Number Variation, Microbiome, Metabolomics, Proteomics, Galaxy, 3D-BioInformatics, Intrinsically Disordered Proteins, Microbial Biotechnology and Plant Sciences. The Communities drive the work of the ELIXIR Platforms by describing their bioinformatics requirements. A close collaboration between the ELIXIR Communities and Platforms safeguards services developed by the ELIXIR Platforms would be fit for purpose.

F&N activities are well represented within Europe and ELIXIR nodes. Following the establishment of (inter) national efforts such as ENPADASI, RICHFIELDS, NuGO and the formation of the FNH-RI, an ELIXIR Community on F&N can have a positive impact on the community, strengthening collaboration and delivering a more harmonised strategy amongst service providers. F&N aligns with the ELIXIR Platforms and Communities as well as other health related themes represented and proposed in ELIXIR like human data and rare disease and including omics data. Although the use case (understanding the relation between diet, microbiome, metabolome and health) described above is not the only activity of interest within the ELIXIR Nodes, we proposed this use case as a starting point of common interest to catalyze the collaboration amongst ELIXIR Nodes and ELIXIR Platforms.

## Alignment with ELIXIR Platforms

The F&N Community is focussing on why people eat what they eat, and how that affects their health. Food (sustainable, healthy, affordable, reliable, and preferable), behaviour (purchase, preparation and consumption) and dietary intake data and integration tools are needed to fulfil the current challenges. Moreover, the F&N Community needs to connect to other (biomedical) health-related communities and ESFRIs (BBMRI, ELIXIR and ECRIN) as prevention of disease requires understanding the nature of health relative to disease. The community has developed and will develop big and linked open data solutions and tools as for example, food purchase apps and e-Health, but also models regarding sustainable food and food security (making sure the whole system provides enough food on a sustainable basis) are essential to come to this goal, these may be relevant to ELIXIR. The molecular resources and bioinformatics tools available in ELIXIR are relevant to the F&N Community. Moreover, shared data is available in the community, including several omics related datasets (e.g. transcriptomics, metabolomics and microbiome). These resources are highly relevant to ELIXIR. The metabolomics use case is specifically relevant for the Food and Nutrition community as the effects of food are small and diverse, which makes metabolomics an important platform.

The F&N use case requires all the Platforms of ELIXIR (Tools, Interoperability, Data, Compute, and Training) as all these areas are relevant to the F&N Community. Our well-structured and publicly available databases containing data on healthy 'subjects' (citizens) and composition of foods should become interoperable and part of the ELIXIR FAIR data resources. These data on lifestyle and health prevention will be a new asset to the FAIR data backbone, as there currently is a predominance of patient data. With these datasets FAIRified and interoperable, the ELIXIR tool set will make it possible to analyse data in the whole range from health prevention to disease. Moreover, the different projects and consortia have delivered several databases, software tools and training materials that are useful for a broader community, which is described below. The F&N Community will help to strongly advocate the ELIXIR services and broaden their user base.

## Data Platform

The ELIXIR request for 2021 [Data Platform priorities](#) focuses on sustaining Europe's life science data infrastructures in the long term by working on guidelines and indicators to improve the impact of data resources and their long-term sustainability. Additionally, this platform aims to improve links between curated and non-curated data resources and literature.

The F&N Community realised early on that they need a study capturing tool that can capture the full study design, large study outcome data sets from for instance transcriptomics studies and rich phenotype descriptions and that also offers an analysis platform for integrated data analysis for related studies. The [Phenotype database](#) was developed for this purpose after evaluating other study capture environments available at that time. These alternatives, especially ISA-creator, were not able to capture study designs typical for the domain (cross-over studies) and did not have the necessary support for food intake registration and the nutritional phenotype in general. The continued development of the phenotype database

was supported in the ENPADASI project which also allowed development of a food and nutrition ontology that can be used to capture most aspects that are specific to the field.<sup>14</sup> Further development of that ontology will happen in the new Food Nutrition-Security Cloud project (FNS-Cloud).

Study capture databases like the phenotype database basically are project level databases that come in between the capturing of individual studies happening in the departments, clinics, metabolic wards, and laboratory environments and the scale of technology specific to ELIXIR recommended data repositories and Biosamples/Biostudies. Other examples are Molgenis (used in the rare disease field), ISA-tools and FAIRdom. We think there are important challenges in the integration of study level data capturing. Discussions with especially Molgenis, have indicated that co-development of a template system and software libraries that support ontology term selection based on combinations of free text entry and pull-down menus would be really useful and would in practice lead to collection of more interoperable data. Increased data interoperability and the ability to find and access data between instances of study capture databases would also allow federated analysis across such instances.

### Tools Platform

The Tools Platform drives access and utilisation of bioinformatics research software by working closely with services and connectors. Additionally, this Platform aims to facilitate the discovery (*bio.tools*, EDAM), benchmarking (OpenEBench) and interoperability of bioinformatics software, by focusing on software development best practices (4OSS), and on strategy for workflows and software containers (BioContainers).

The tools provided and used by the F&N Community will be registered in the *bio.tools* registry, and related concepts will be added to EDAM if needed. Training in software best practices and development of Software Management Plans will help the community with providing high-quality computational tools.

### Interoperability Platform

The Interoperability Platform provides support to the discovery, integration and analysis of biological data based on FAIR principles. A set of recommended tools and services have been selected by the Interoperability platform and they are named Recommended Interoperability Resources, including persistent identifiers, metadata, and data markups (Bioschemas), standard for workflow description, registries for ontologies, controlled vocabularies, exchange and storage formats. This Platform facilitates work on the description of interoperability services and organises specialised BYOD (Bring Your Own Data) workshops with the aim to improve the FAIRness of data resources.

F&N data must be interoperable in order to align with other data sources to be able to answer the complex scientific challenges. Extension of the current metadata standards toward food and consumer science is needed to make this possible. Part of this has been included in ENPADASI and Richfields, but full integration with other interoperability platforms is lacking and black areas are still present.

### Compute Platform

The Compute Platform is devoted to the compute, transfer, storage, authentication, and authorization related to biological data relying on services provided by ELIXIR Nodes and other e-infrastructures.

Key elements of the Compute Platform are:

1. Identity and access management, including AAI.
2. Making datasets available in relevant cloud providers.
3. Defining and coordinating an ELIXIR hybrid cloud ecosystem.
4. Community containers being deployed and operated at scale.

It doesn't directly provide resources but can help in brokering access.

The interaction with the compute platform will allow the F&N Community to process an increasing number of relevant experimental data (i.e. omics) using standardized bioinformatics pipelines and statistical analyses. Therefore, the resulting data structure will be suitable for comparison, integration and modelling. The computing power necessary for the analyses that the F&N Community needs for the development and application of nutrition health care models will be huge. The Community should have a helpdesk that knows the access points for the Compute Platforms and how to use them.

## Training Platform

Training is a key component of the sustainability of a community. We will make all training material available for the ELIXIR Community to collect data on food composition, dietary assessment, and food related behaviour of highest scientific quality to support food intake studies in surveillance, nutritional interventions and clinical studies. For new food and health web-based and mobile applications new training material will be developed (based on a structured gap analysis on the training needs). Training on the use of measuring behaviour, dietary intake, food matching and imaging algorithms and tools, and their limitations will be made accessible. Moreover, the training packages developed by ENPADASI on nutritional data upload and by RICHFIELDS and ENPADSI on guidelines in relation to ethics, privacy and IP with a focus on data sharing will be made part of the ELIXIR training portfolio.

Courses from the F&N Community will be registered in **TeSS** (the ELIXIR Training portal) and first steps will be taken to define learning paths considering the different starting points of the users. On the other hand, ELIXIR courses, such as those on tools and services related to data management and stewardship will be necessary to the F&N Community. Special attention, also due to COVID-19 challenges, should be focused on e-learning materials, tools and services and virtual or hybrid training/CB events. ELIXIR training platform services are listed on the **Platform's website**.

### *Connection with innovation and SME forum*

Omic approaches are becoming a reality also at the company level and not only in the academic field. A rapid and cost-effective and on-site technology represents the gateway to the application of (nutri-)genomics directly in food production and nutritional screening supporting technology transfer. This has led to a positive increase to data production and access by private citizens (e.g. microbiome screening) but also sometimes an incorrect use of analytical techniques, with a potential long-term impact on data reliability. The guidelines for correct data and pipeline integration may derive from the connection between the F&N Community and the SME forum in order to collect the real needs of companies, in terms of training, capacity building and research lines.”

Alignment with ELIXIR communities and focus groups

The ELIXIR Platforms are currently complemented by 11 scientific Communities, here we indicate the Communities and focus groups that are most relevant to the F&N community:

- The Federated Human Data community for long-term strategies for managing and accessing sensitive human data and connecting consumer and patient data.
- The Rare Diseases community for privacy issues on the individual data and describing phenotypes.
- The Marine Metagenomics (Microbiome) community for the solutions in the area of microbiome/metagenome analysis.
- The Biodiversity Focus Group for the accessibility to taxonomic and molecular data (including other metadata) related to the species described so far (biodiversity catalogues).
- Plant Science Community for the link between plant science in general and plants as food compounds.
- The Metabolomics Community for readouts of intake and health.
- The Toxicology (not yet an approved Community) on describing phenotypes.
- The newly developing Microbiome community and the Microbial Biotechnology Community for two different microbiome approaches.
- The Machine learning Focus Group for complex data integration (including omics and personalization) with specific focus on metabolomics and microbiome, and software and pipelines for analysis.
- Other Communities are Proteomics, Galaxy, 3D-Bioinformatics and Intrinsically Disordered Proteins.

The F&N Community is in several ways in line with the other communities defined so far in ELIXIR and requires similar data solutions, but it also adds new data sources and new solutions to the current ELIXIR Communities. For instance, the

F&N Community adds a consumer perspective to the current Communities in ELIXIR. Working with consumer data has privacy and ethical issues that are in line with the other communities (especially the rare diseases because of the individual data and the possibility of de-anonymization). Interestingly people also collect health and food-intake related data with smart phones and other personal devices, including sensors to capture non or minimally invasive biometric signals in real time. Individual digital health is becoming a revolution. Advances in sensor design, smart device connectivity and data acquisition help to keep track of parameters such as food intake, calories burnt or activity levels to complement self-reporting dietary electronic notebook.<sup>41</sup> In addition, smart devices can be used to send meal photographs, notifications and reminders to complete tasks providing additional data to evaluate diet adherence or consumer experience.<sup>7</sup> This is expected to become more relevant and especially more used in the F&N field rapidly. For ELIXIR this means involvement in a new and modern data field that wasn't covered so much up to now. The individual data collection needed for this is in line with the work of the Rare Diseases Community.

The current Federated Human Data Community is directed towards patients, which could create mutual enrichments by aligning with the data on healthy free-living individuals that are collected in the F&N Community. To make this possible, alignment of standards is needed.

The relation between dietary intake, microbiome and health is becoming more and more obvious. Standardization of metagenomics research is needed to bring the research to the next level. The Marine Metagenomics (Microbiome) Community is actively working in the area and the F&N Community can benefit from their developments. Links to workflows, sequence reference resources and tools for taxonomic/genetic profiling of microbiomes developed by the Microbiome Community are relevant to the F&N Community. For example, ELIXIR-Italy provided amplicons and shotgun metagenomic data analysis tools, DNA barcoding reference databases and recently a virtual research environment targeted in particular at eukaryotic microbial communities. In the context of this connection, specific sections in the described IT resources could be dedicated to nutrition-related microorganisms (such as gut microbes, food production chain microbes, probiotics...). In this respect links to the Biodiversity Focus Group and the Microbiome Community are relevant as well. For example, the increasing connection of nutritional -omics datasets will contribute to the identification of unbiased microbial biomarkers. This is difficult, given the intrinsic nature of the microbiome and its genetic pool. In contrast to the human inherited, largely static and "non-coding" genome, the metagenome is gained and has a very dynamical composition in response to a multitude of factors. Often viewed as hosting only pathogens, causing disease, or as passive bystanders, it is now observed that the microbiome has a wide 'grey zone' that cannot be simply classified into this dichotomy. Many 'pathogens' inhabit disease free-hosts (e.g., *Helicobacter pylori*) and also commensals may promote pathology onset under certain conditions (e.g. immuno depressed or highly stressed subjects). Because of this high variability between individuals, it is very difficult to establish what the normal condition is, what dysbiosis is and whether the latter is associated with an unhealthy phenotype<sup>42</sup> or diet. For this, the F&N Community has proposed some general metrics, such as alpha species diversity, the ratio of Firmicutes to Bacteroidetes phyla, and the relative abundance of beneficial genera versus facultative anaerobes or pro-inflammatory microbes.<sup>43</sup>

The role of the microbiome also spans beyond the microbiome composition and requires functional analysis. A central question is whether we can predict and model the metabolic activities in the gut starting from available, mostly sequence, data. In a practical sense that leads to evaluation of DNA and RNA sequences to predict the presence of enzymatic activity on the protein level using recognition of translated domains and mapping to microbiome genome scale metabolic models and combination with metabolomics data. This has clear links with the Microbial Biotechnology and Metabolomics Communities and for the complex integration of these omics data and the personalization of nutritional interventions with the machine learning community and also to the currently active Systems Biology Focus Group.

Understanding the effect of food on health requires measures of intake and of health; metabolomics platforms can deliver these markers as has been shown in the JPI HDHL project, Foodball. Therefore, a link to the Metabolomics Community is very important to the F&N Community. Linking to other communities is also important; for example, linking to the plant science community is crucial because of the interrelationship between plant nutrients and their effects on human metabolism and health. Large-scale transcriptomics studies demonstrate that edible plants, also including their derivatives such as olive oil, can induce significant changes in the human gene expression profile including many ncRNAs (i.e. miRNAs and lncRNAs) that are key regulators of gene expression. It is noteworthy that plant nutrients have been demonstrated to have a positive impact on many signalling and metabolic pathways related to diabetes, obesity, neurodegenerative disorders and in general to the immune response to stress factors and inflammation. At cell process level plant nutrients have been demonstrated to have positive effects (i.e. gene expression) on different DNA repair mechanisms, apoptosis, oxidative phosphorylation and mitochondrial metabolism, just to cite only the most relevant. To elucidate the molecular mechanisms and the plant nutrients able to induce such beneficial effects on human health requires an effective integration of the Plant Science, Metabolomics and F&N Communities at level of competences, data resources and analysis tools. The F&N Community members have partly implemented the FAIR principles in

ENPADASI and have developed ontologies, quality assessment tools, and standardized questionnaires that have been made publicly available as much as possible.

The F&N Community brings data on consumer science and ways to collect them to ELIXIR. Moreover, we bring ways to collect rich meta-data e.g. on cross-over designs, healthy phenotypes, challenge studies, protocol sets, dietary data and dietary questionnaires.

### Alignment with other ESFRIs

The F&N community is aligned with several other ESFRI; see [Table 1](#) for these interactions.

### Related projects

A number of projects have previously developed bioinformatics tools for the Food & Nutrition field. These may also become for the ELIXIR Food & Nutrition Community. Some of these projects are producing data and/or tools which can be used by the F&N Community (see [Table 2](#)).

**Table 1. The alignment of ELIXIR's Food & Nutrition Community with other ESFRI's focus.**

ESFRI	Alignment with ELIXIRs Food & Nutrition Community
FNH-RI (and EuroDISH)	Food, nutrition and health RI is a research infrastructure under development. The described community is the same as the F&N. Therefore the description in this article also reflects how FNH-RI, if established, will collaborate with ELIXIR.
BBMRI	Has important activities around patient registries and capturing and analyzing patient study data in the Molgenis environment. This has relevant parallel developments which we have identified in shared meetings in the past and which lead to the exploration of integration of shared templates and even exchange modules of the dbNP and Molgenis platform. The two interaction of the F&N community with BBMRI can also benefit from the difference in focus "health" versus "disease", modern medicine also moves towards more individual and health targeted approaches and can benefit from health directed data capturing developed in the F&N environment (e.g. stress tests, and personal health monitoring via apps), while the F&N community can benefit from more structured descriptions of clinical conditions and measurements developed in BBMRI.
EATRIS	Has its focus on translational research and thereby industry collaboration and has set up many instruments to support public-private partnerships which will also be useful when we want to develop new collaborations with the agro-food, retailer and health monitoring sectors. This includes models for data sharing and project sustainability. EATRIS also has experience with EU Horizon related partnerships that can be developed between amongst others commercial sectors and the EU. Examples like IMI (drug related research with EFPIA companies) and Colipa (cosmetics companies often funding research together) can possibly serve as models for collaboration with F&N and health related companies.

**Table 2. Projects of relevance to the ELIXIR Food & Nutrition Community.**

Project Name	Project Website
NuGO	<a href="http://www.nugo.org/">http://www.nugo.org/</a>
MPG	<a href="https://www.wikipathways.org/index.php/Portal:Micronutrient">https://www.wikipathways.org/index.php/Portal:Micronutrient</a>
ENPADASI	<a href="http://www.enpadasi.eu/">http://www.enpadasi.eu/</a>
EuroFIR	<a href="https://www.eurofir.org">https://www.eurofir.org</a>
FNS-Cloud	<a href="https://www.fns-cloud.eu/">https://www.fns-cloud.eu/</a>
Food4Me	<a href="https://www.food4me.org">https://www.food4me.org</a>
FoodBall	<a href="https://foodmetabolome.org">https://foodmetabolome.org</a>
FoodPhyt	<a href="https://research.chalmers.se/en/project/9396">https://research.chalmers.se/en/project/9396</a>
QuaLiFY	<a href="http://www.eurodish.eu/">http://www.eurodish.eu/</a>
Environment task in EJP-RD	<a href="https://www.ejprarediseases.org/">https://www.ejprarediseases.org/</a> identifying role of (micro) nutrient in rare disease networks
CIRCLES	<a href="https://circlesproject.eu/">https://circlesproject.eu/</a> Controlling mICRobiomes CircuLations for bETter food Systems
MASTER	Sustainable Food Security

**Table 3. Overview of tools and data available to the Food & Nutrition Community and short explanations highlighting the benefits.**

Tools/Data	Short explanation and benefit for ELIXIR F&N Community
SOP portal	Procedure sharing and analysis provenance
NuGOWiki	Annotating food molecular food constituents with nutritional information
NuGO Blackbox	Needs for automatically updating, shared and collaboratively maintained compute infrastructure based on BioLinux. A kind of early cloud. But also an early collection of needed tools and ideas for NuGO single sign-on (now available as ELIXIR AAI)
Micronutrient Genomics Project	Expert groups for different groups of micronutrients; did collaborative research on that and started to create micronutrient pathways, with its own portal on WikiPathways, and to prepare reviews for individual micronutrients, network(?) evaluation for combinations of effects (e.g. in epigenetics and in antioxidant effects) Possibility to move that to projects, although only funded as an IRSES for exchanges.
Phenotype Database	DASH-IN, and ENPADASI project
FoodBALL resources	<a href="http://foodmetabolome.org/">http://foodmetabolome.org/</a>
Nutritional Ontology	ONE and ONS
NuGOarray	For legacy only, not likely to lead to new needs
Biomarker development	Including challenge tests for health biomarkers (discussed above)
Nutritional systems biology	Typically changes in diet act in concert and affect many aspects of the system, so it is less "find the interesting, affected gene/protein or possible drug target" and more "understand the small system changes that act in concert". That meant that nutritional research was amongst the first to adopt technologies aiming for such system wide effects like pathway and network analysis.
NutrigenomeDB	Easy-to-use web application that allows exploration of differential gene expression profiles from nutrigenomics experiments through data tables and interactive visualizations
Food intake	Food4Me Food Frequency Questionnaire (FFQ); validated electronic measurement of food intake, available in 8 languages & food images for estimation of amount of food
	Foodbook24; validated web based dietary assessment based on 24 h recall
	Food Profiler; validated App in multiple EU countries for collection of user data on food consumption patterns and consumer behaviour.
	Healthy Eating Index calculator; Healthy eating index can be calculated based on food groups derived from Food4Me FFQ and is a measure for diet quality according to EU standards
Knowledge rules	Food4Me decision trees; a decision support tool that uses a tree-like graph or model of decisions that uses biological user data (intake, genetics, health, anthropometrics) and delivers nutritional recommendations based on this user data.
	Metabotypes decision trees; a decision support tool that uses a tree-like graph or model of decisions based on subgroups of users that matches on their clinical chemistry and anthropometric data and receive dietary advice tailored to this 'metabotype'.
Food composition	EuroFIR FoodExplorer™ & FoodBasket™; web based delivery of information on food composition and recipes

**Table 3.** *Continued*

<b>Tools/Data</b>	<b>Short explanation and benefit for ELIXIR F&amp;N Community</b>
Food composition, food consumption, Total Diet Study and brands	FoodCASE, an information system to manage and generate food composition, food consumption, total diet study and brand data. The tool offers several functionalities to estimate, analyse, link, visualise, and publish data.
Rules toolset	A cloud-based, efficient tool for delivering personalised healthcare information. The Rules Toolset is a unique product that supports scientists to transform the synergistic input of nutritional, biological, medical and genetic information into a comprehensive report in the simplest way, regardless of the complexity of the logic.
RICHFIELDS architecture	Platform with food, health, behaviour reference data and ontologies, and knowledge rules
Food classification & description tools	Languag Food product Indexer; FoodEX2 Browser Tool
ECRIN Nutrition ontology	Taxonomy and ontology
EFSA Food ontology	Taxonomy and ontology
RICHFIELDS Behaviour ontology	Taxonomy and ontology
Integrated web services	APIs to food reference databases and knowledge rules
Smart Food Intake	2 hr dietary recall and underlying food choice motivations
mobile Food Record	Image-based intake assessment tool (Nestlé)
Automated Consumer-centred advice generation	knowledge rules based on Bayesian Belief Networks integrating with drivers of food choice & psychosocial factors
3D printer	Personalised food manufacturing
Smart food manufacturing hub	Personalised food manufacturing
Personalised menu Planners	Via linear modelling approach (Qualify)



## Data and tools developed by the F&N community

Table 3 shows an initial overview of data and tools that the F&N community has available at the time being. During the upcoming Implementation Study this will be extended.

## Data availability

No data is associated with this article.

## References

- Pechey R, Monsivais P: **Socioeconomic inequalities in the healthiness of food choices: Exploring the contributions of food expenditures.** *Preventive Medicine.* 2016 Jul 1; **88**: 203–209. [Publisher Full Text](#)
- Leeming ER, Johnson AJ, Spector TD, et al.: **Effect of diet on the gut microbiota: rethinking intervention duration.** *Nutrients.* 2019 Dec; **11**(12): 2862. [Publisher Full Text](#)
- Dabke K, Hendrick G, Devkota S: **The gut microbiome and metabolic syndrome.** *The Journal of Clinical Investigation.* 2019 Oct 1; **129**(10): 4050–4057. [Publisher Full Text](#)
- Munoz-Garach A, Diaz-Perdigones C, Tinahones FJ: **Gut microbiota and type 2 diabetes mellitus.** *Endocrinología y Nutrición (English Edition).* 2016 Dec 1; **63**(10): 560–568. [Publisher Full Text](#)
- Rajani C, Jia W: **Disruptions in gut microbial-host co-metabolism and the development of metabolic disorders.** *Clinical Science.* 2018 Apr 16; **132**(7): 791–811. [Publisher Full Text](#)
- Halmos T, Suba I: **Physiological patterns of intestinal microbiota. The role of dysbacteriosis in obesity, insulin resistance, diabetes and metabolic syndrome.** *Orvosi Hetilap.* 2016 Jan 1; **157**(1): 13–22. [Publisher Full Text](#)
- Berry SE, Valdes AM, Drew DA, et al.: **Human postprandial responses to food and potential for precision nutrition.** *Nature Medicine.* 2020 Jun; **26**(6): 964–973. [Publisher Full Text](#)
- Korem T, Zeevi D, Zmora N, et al.: **Bread affects clinical parameters and induces gut microbiome-associated personal glycaemic responses.** *Cell Metabolism.* 2017 Jun 6; **25**(6): 1243–1253.e5. [Publisher Full Text](#)
- Lamichhane S, Sen P, Dickens AM, et al.: **Gut metabolome meets microbiome: A methodological perspective to understand the relationship between host and microbe.** *Methods.* 2018 Oct 1; **149**: 3–12. [Publisher Full Text](#)
- Nichols RG, Peters JM, Patterson AD: **Interplay between the host, the human microbiome, and drug metabolism.** *Human Genomics.* 2019 Dec; **13**(1): 1–0. [Publisher Full Text](#)
- Thaiss CA, Levy M, Korem T, et al.: **Microbiota diurnal rhythmicity programs host transcriptome oscillations.** *Cell.* 2016 Dec 1; **167**(6): 1495–1510.e12. [Publisher Full Text](#)
- Debeljak J, Ferik P, Čokolič M, et al.: **Randomised, double blind, cross-over, placebo and active controlled human pharmacodynamic study on the influence of silver fir wood extract (Belinal) on post-prandial glycaemic response.** *Die Pharmazie-An International Journal of Pharmaceutical Sciences.* 2016 Oct 1; **71**(10): 566–569.
- Baldani DP, Skrgatic L, Cerne JZ, et al.: **Association of PPARG Pro12Ala polymorphism with insulin sensitivity and body mass index in patients with polycystic ovary syndrome.** *Biomedical Reports.* 2014 Mar 1; **2**(2): 199–206. [Publisher Full Text](#)
- Vitali F, Lombardo R, Rivero D, et al.: **ONS: an ontology for a standardized description of interventions and observational studies in nutrition.** *Genes & Nutrition.* 2018 Dec; **13**(1): 1–9. [Publisher Full Text](#)
- Macdiarmid J, Blundell J: **Assessing dietary intake: who, what and why of under-reporting.** *Nutrition Research Reviews.* 1998 Dec; **11**(2): 231–253. [Publisher Full Text](#)
- Pfimer K, Vilela M, Resende CM, et al.: **Under-reporting of food intake and body fatness in independent older people: a doubly labelled water study.** *Age and Ageing.* 2014 Oct 22; **44**(1): 103–108. [Publisher Full Text](#)
- Scalbert A, Brennan L, Manach C, et al.: **The food metabolome: a window over dietary exposure.** *The American Journal of Clinical Nutrition.* 2014 Jun 1; **99**(6): 1286–1308. [Publisher Full Text](#)
- Maruvada P, Lampe JW, Wishart DS, et al.: **Perspective: dietary biomarkers of intake and exposure—exploration with omics approaches.** *Advances in Nutrition.* 2020 Mar 1; **11**(2): 200–215.
- Gao Q, Praticò G, Scalbert A, et al.: **A scheme for a flexible classification of dietary and health biomarkers.** *Genes & Nutrition.* 2017 Dec; **12**(1): 1–5. [Publisher Full Text](#)
- Praticò G, Gao Q, Scalbert A, et al.: **Guidelines for Biomarker of Food Intake Reviews (BFIRev): how to conduct an extensive literature search for biomarker of food intake discovery.** *Genes & Nutrition.* 2018 Dec; **13**(1): 1–4. [Publisher Full Text](#)
- Dragsted LO, Gao Q, Scalbert A, et al.: **Validation of biomarkers of food intake—critical assessment of candidate biomarkers.** *Genes & Nutrition.* 2018 Dec; **13**(1): 1–4. [Publisher Full Text](#)
- Dragsted LO, Gao Q, Praticò G, et al.: **Dietary and health biomarkers—time for an update.** *Genes & Nutrition.* 2017 Dec; **12**(1): 1–7. [Publisher Full Text](#)
- Giesbertz P, Brandl B, Lee YM, et al.: **Specificity, Dose Dependency, and Kinetics of Markers of Chicken and Beef Intake Using Targeted Quantitative LC-MS/MS: A Human Intervention Trial.** *Molecular Nutrition & Food Research.* 2020 Mar; **64**(5): 1900921. [Publisher Full Text](#)
- Cuparencu C, Rinnan Å, Silvestre MP, et al.: **The anserine to carnosine ratio: an excellent discriminator between white and red meats consumed by free-living overweight participants of the PREVIEW study.** *European Journal of Nutrition.* 2021 Feb; **60**(1): 179–192. [Publisher Full Text](#)
- Anesi A, Mena P, Bub A, et al.: **Quantification of urinary phenyl-γ-valerolactones and related val.**
- Vázquez-Manjarrez N, Weinert CH, Ulaszewska MM, et al.: **Discovery and validation of banana intake biomarkers using untargeted metabolomics in human intervention and cross-sectional studies.** *The Journal of Nutrition.* 2019 Oct 1; **149**(10): 1701–13. [Publisher Full Text](#)
- Gao Q, Dragsted LO, Ebbels T: **Comparison of Bi- and Tri-Linear PLS Models for Variable Selection in Metabolomic Time-Series Experiments.** *Metabolites.* 2019 May; **9**(5): 92. [Publisher Full Text](#)
- SC Sri Harsha P, Abdul Wahab R, Cuparencu C, et al.: **A metabolomics approach to the identification of urinary biomarkers of pea intake.** *Nutrients.* 2018 Dec; **10**(12): 1911. [Publisher Full Text](#)
- Brouwer-Brolsma EM, Brennan L, Devon CA, et al.: **Combining traditional dietary assessment methods with novel metabolomics techniques: present efforts by the Food Biomarker Alliance.** *Proceedings of the Nutrition Society.* 2017 Nov; **76**(4): 619–627. [Publisher Full Text](#)
- Cuparencu C, Rinnan Å, Dragsted LO: **Combined markers to assess meat intake—human metabolomic studies of discovery and validation.** *Molecular Nutrition & Food Research.* 2019 Sep; **63**(17):

1900106.  
[Publisher Full Text](#)
31. Andersen MB, Rinnan Å, Manach C, *et al.*: **Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern.** *Journal of Proteome Research*. 2014 Mar 7; **13**(3): 1405–1418.  
[Publisher Full Text](#)
  32. Acar E, Gürdeniz G, Khakimov B, *et al.*: **Biomarkers of Individual Foods, and Separation of Diets Using Untargeted LC-MS-based Plasma Metabolomics in a Randomized Controlled Trial.** *Molecular Nutrition & Food Research*. 2019 Jan; **63**(1): 1800215.  
[Publisher Full Text](#)
  33. Meslier V, Laiola M, Roager HM, *et al.*: **Mediterranean diet intervention in overweight and obese subjects lowers plasma cholesterol and causes changes in the gut microbiome and metabolome independently of energy intake.** *Gut*. 2020 Jul 1; **69**(7): 1258–1268.  
[Publisher Full Text](#)
  34. Lachat C, Hawwash D, Ocké MC, *et al.*: **Strengthening the Reporting of Observational Studies in Epidemiology–nutritional epidemiology (STROBE-nut): An extension of the STROBE statement.** *Nutrition Bulletin*. 2016 Sep; **41**(3): 240–251.  
[Publisher Full Text](#)
  35. Yang C, Hawwash D, De Baets B, *et al.*: **Perspective: Towards Automated Tracking of Content and Evidence Appraisal of Nutrition Research.** *Advances in Nutrition*. 2020 Sep 1; **11**(5): 1079–1088.  
[Publisher Full Text](#)
  36. Eftimov T, Korošec P, Koroušić SB: **StandFood: standardization of foods using a semi-automatic system for classifying and describing foods according to FoodEx2.** *Nutrients*. 2017 Jun; **9**(6): 542.  
[Publisher Full Text](#)
  37. Popovski G, Seljak BK, Eftimov T: **A survey of named-entity recognition methods for food information extraction.** *IEEE Access*. 2020 Feb; **8**(8): 31586–31594.  
[Publisher Full Text](#)
  38. Popovski G, Korousic-Seljak B, Eftimov T: **FoodOntoMap: Linking Food Concepts across Different Food Ontologies.** *In KEOD*. 2019; **2019**: 195–202.  
[Publisher Full Text](#)
  39. Popovski G, Seljak BK, Eftimov T: **FoodBase corpus: a new resource of annotated food entities.** *Database*. 2019 Jan 1; **2019**: 2019.  
[Publisher Full Text](#)
  40. Stojanov R, Popovski G, Jofce N, *et al.*: **Foodviz: Visualization of food entities linked across different standards.** *In International Conference on Machine Learning, Optimization, and Data Science*. Cham: Springer; 2020 Jul 19; (pp. 28–38).  
[Publisher Full Text](#)
  41. Mohammed MS, Sendra S, Lloret J, *et al.*: **Systems and WBANs for controlling obesity.** *Journal of Healthcare Engineering*. 2018 Feb 1; **2018**: 1–21.  
[Publisher Full Text](#)
  42. Scott AJ, Alexander JL, Merrifield CA, *et al.*: **International Cancer Microbiome Consortium consensus statement on the role of the human microbiome in carcinogenesis.** *Gut*. 2019 Sep 1; **68**(9): 1624–1632.  
[Publisher Full Text](#)
  43. Hills RD, Pontefract BA, Mishcon HR, *et al.*: **Gut microbiome: profound implications for diet and disease.** *Nutrients*. 2019 Jul; **11**(7): 1613.  
[Publisher Full Text](#)

# Open Peer Review

Current Peer Review Status: ? ?

---

## Version 1

Reviewer Report 09 October 2023

<https://doi.org/10.5256/f1000research.54943.r211055>

© 2023 Bansal A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

? **Anu Bansal** 

Lovely Professional University, Phagwara, Punjab, India

The manuscript provides a thorough review of the challenges encountered in the field of food & nutrition highlighting the necessity of interdisciplinary cooperation and data integration to address these issues effectively. The authors focused on the significance of consistent data collection, sharing, and analysis to improve reader's knowledge of dietary decisions and their effects on health. With some minor revisions to improve clarity, this manuscript has the potential to make a significant contribution to the field.

Suggestions for improvement:

1. Although the manuscript offers a plethora of material, its organization and clarity may be enhanced. Some readers may find it difficult to follow the subject matter because of how intricate it is. For a better reading, think about dividing some of the larger paragraphs.
2. In-text citations: The format of citations should be the same throughout the manuscript. See line: "However, to date, efforts are still fragmented, and the integration requires a wide range of data science [1], and better ontologies...".
3. Careful grammar and language editing is advised.
4. The manuscript might use some case studies or real examples to demonstrate the practical difficulties in integrating F&N data.
5. Although the alignment with the ELIXIR Communities and Focus Groups is extensively explored, it would be beneficial to include specific suggestions or tactics for how these alignments could be put into practice.

## References

1. Jan I, Sharma P, Bansal P: Microbial Bioinformatics Approach in Food Science. In: Karnwal, A., Mohammad Said Al-Tawaha, A.R. (eds) Food Microbial Sustainability. *Springer, Singapore*. 2023.

267-288 [Reference Source](#)

**Is the topic of the opinion article discussed accurately in the context of the current literature?**

Yes

**Are all factual statements correct and adequately supported by citations?**

Partly

**Are arguments sufficiently supported by evidence from the published literature?**

Yes

**Are the conclusions drawn balanced and justified on the basis of the presented arguments?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Biochemistry, Phytochemistry, Computational studies, and Bioinformatics.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Reviewer Report 20 September 2022

<https://doi.org/10.5256/f1000research.54943.r148525>

© 2022 King G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Graham King**

Southern Cross Plant Science, Southern Cross University, Lismore, NSW, Australia

This opinion article summarises interactions from a workshop and the potential role for ELIXIR. It provides a useful and generally balanced overview of the issues and current constraints on data availability and interchange in the field of food and nutrition. It first outlines the rationale for personalised nutrition and highlights the complexity of data and relations that are associated with food.

I found the overall presentation patchy – a mix of structured headings and bullet-point lists. I would suggest these are cleaned up and the bullet-point lists presented in tables, especially for those describing the workshop challenges, which may then be referred to in subsequent sections that elaborate on these points.

Table 3 is useful, but could be enhanced and structured further – or cross-refer to a network

knowledge graph figure. Even more value would be added were it possible to include columns with a critical assessment of each of F, A, I and R for each resource.

Although the need for more data in this expanding field is correctly stated at the outset, this needs to be balanced by a thorough global community-wide discussion of common and overlapping use of language and meta-data. The field of food and nutrition is marked by multiple independent resources, for which some degree of interoperability is starting to emerge for a subset. Thus, the gap is not 'just more data' but the need for data to be structured and annotated in a consistent manner – e.g. through consensus establishment and development of relevant overlapping ontologies representing relevant domains.

This aspect is covered further on in the article, but perhaps should be clear from the outset, as is particularly important to facilitate the trans-disciplinary approaches from food production to consumption.

Based on the information presented, there appears to be a legitimate role for the ELIXIR network to provide a focus for interoperable platforms that may avoid some of the locked-in pitfalls of system evolution vs well considered and structured information engineering design. Community wide consultation is essential for such efforts, especially from representatives who are not hard-core informaticians. The headings and content of the article will hopefully help in development of a roadmap, not only in relation to ELIXIR, but for ELIXIR to interact with other infrastructure providers worldwide.

Given the broad scope and context it is not surprising there are some omissions, with primarily European-centric viewpoints presented. I suggest that if possible the authors ensure they widen the scope and highlight where there is opportunity for integration with initiatives in N. America and elsewhere. This is noted in some places but should be considered throughout.

Another gap appears in the consideration of food composition – in particular, there is little if any consideration of the sources of *variation* in nutritional composition (i.e. food primary production, supply and processing) and how data on food composition should be formalized and structured in meaningful way. One would expect interaction with Initiatives such as FoodON and CDNO within the OBO Foundry. Nutrition-related ontologies such as ONS and ONE are mentioned, but currently are limited in scope and do not provide sufficient depth to associate human health with dietary nutritional composition. It should also be noted that some of the significant food composition data sources mentioned are proprietary and currently do not meet FAIR standards.

Static food composition tables are a proxy for the reality of what is consumed – if personalised human health is to become a reality this must be matched by the reality of personalised dietary intake. e.g. in terms of supply chain/seasonal/batch variation that takes into account the influence of production environment, and is particularly relevant to fresh food and associated phytonutrients.

Although many ontologies are available to describe entities and processes in the food and nutrition field – one would hope for some critical evaluation in this article of their relative worth and availability of use-cases. The authors recognize that development of ESFRI requires better ontologies and dbs etc., and hopefully a sector-wide consideration of how ontologies may better re-use common terms etc. It would be useful to adopt some formalized quality criteria to assess

the information content, completeness and uniqueness of relevant ontologies mentioned – do they all adhere to OBO principles, are they actively maintained and engage in open (eg GitHub) discussion with global community? Are they structured in an approachable and meaningful way for the domain?

Finally, in describing the alignment with the ELIXIR platform – I suggest a strong link be described between Ag (production) and Health. It is note-worthy that the article is included in the F1000 'Agriculture, Food & Nutrition' gateway, yet no mention of the source of variation in food.

*Additional minor points:*

I found some of the language to be awkward. e.g. first sentence of section 'individual health concept'.

How are issues of personal privacy to be managed in the future (retailers already breach this for food consumption habits)?

Where does the food processing and marketing industry fit into the world-view described in this article? It is noted that there is a connection with innovation and SME forum – again a bullet point list that would better be presented in a table.

The section headed 'F&N community in ELIXIR' did not appear complete or well structured.

The following section on technical activities states "ELIXIR Nodes and supported by the Hub" - this appears to be first mention of Hub in the document and is unclear to what this refers.

**Is the topic of the opinion article discussed accurately in the context of the current literature?**

Yes

**Are all factual statements correct and adequately supported by citations?**

Yes

**Are arguments sufficiently supported by evidence from the published literature?**

Yes

**Are the conclusions drawn balanced and justified on the basis of the presented arguments?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Food production, Ag-nutrition variation, Ontology development and data annotation, Bioinformatics.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**