

Technical Note

The Da Vinci European BioBank: A Metabolomics-Driven Infrastructure

Dario Carotenuto ¹, Claudio Luchinat ^{2,3}, Giordana Marcon ¹, Antonio Rosato ^{2,3}
and Paola Turano ^{1,2,3,*}

¹ Da Vinci European BioBank, FiorGen Foundation, via Sacconi 6, 50019 Sesto Fiorentino, Florence, Italy; E-Mails: dario.carotenuto@gmail.com (D.C.); g.marcon@davincieuropeanbiobank.org (G.M.)

² Magnetic Resonance Center (CERM), University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy; E-Mails: luchinat@cerm.unifi.it (C.L.); rosato@cerm.unifi.it (A.R.)

³ Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

* Author to whom correspondence should be addressed; E-Mail: turano@cerm.unifi.it; Tel.: +39-055-457-4262; Fax: +39-055-457-4926.

Academic Editor: Lori A. Orlando

Received: 8 January 2015 / Accepted: 16 April 2015 / Published: 22 April 2015

Abstract: We present here the organization of the recently-constituted da Vinci European BioBank (daVEB, <https://www.davincieuropeanbiobank.org/it>). The biobank was created as an infrastructure to support the activities of the FiorGen Foundation (<http://www.fiorgen.net/>), a nonprofit organization that promotes research in the field of pharmacogenomics and personalized medicine. The way operating procedures concerning samples and data have been developed at daVEB largely stems from the strong metabolomics connotation of FiorGen and from the involvement of the scientific collaborators of the foundation in international/European projects aimed to tackle the standardization of pre-analytical procedures and the promotion of data standards in metabolomics.

Keywords: biobanking; metabolomics; standard operating procedures; nuclear magnetic resonance; database

1. Introduction

The creation of a biobank infrastructure linking bio-specimens with clinical data generally originates within a clinical context and/or focuses on specific health or disease problems. These infrastructures can generally be reconducted to one of the following formats: Population-based biobanks, disease-oriented biobanks, case-control biobanks, biospecimen specific biobanks (tissues, cord blood, *etc.*), clinical trial related biobanks.

Here, we present the establishment of a methodology-driven biobank, which aims at (although does not limit itself to) the collection of samples and data for future metabolomics analyses and develops protocols and standards that can assure optimal outputs from the use of this analytical methodology. The present case history concerns the creation of the da Vinci European BioBank (daVEB) [1], a small scale infrastructure run by the non-profit Fiorgen Foundation which promotes collaborative interdisciplinary problems in the field of pharmacogenomics and personalized medicine. The centralized IT infrastructure, the laboratory and the first biorepository of daVEB are located at the scientific campus of the University of Florence and hosted by CERM (www.cerm.unifi.it). daVEB is a partner biobank in BBMRI-ERIC (Biobanking and Biomolecular Resources Research Infrastructure-European Research Infrastructure Consortium; www.bbmri-eric.eu) and its Italian hub BBMRI.it. In 2011, daVEB obtained quality certification according to UNI EN ISO 9001:2008 for “Collection, storage and distribution of biological samples and the associated data for scientific research” (CSQ Certificate N. 9122.FFFI; IQNet Certificate N. IT-79101); the certification has been renewed in 2014. Procedures and standards adopted at daVEB originate from the strong links with scientists of the University of Florence and their involvement in several EC projects in related fields.

In this article, after a brief presentation of the key aspects of metabolomics in the frame of biobanks, we will review the actions undertaken at daVEB to develop validated procedures for the collection, handling and storage of samples, to provide adequate data sets for the interpretation of metabolomics data through statistical approaches and, last but not least, to facilitate the distribution of the information deriving from metabolomics analyses within the biomedical research community.

2. Metabolomics

Metabolomics is the study of the complete pool of small molecules forming the metabolome of a biological sample (biofluid, cell, tissue, or organism). The human metabolome composition accounts for several thousands of small molecules (<2000 Da) produced by the genome of the host organism and by the genomes of its microflora or deriving from the interaction with the environment [2]. The concentrations of the different small molecules, generically called metabolites, span a very ample range, from <1 nM to 1 μ M.

The two main analytical methodologies used in metabolomics are mass spectrometry (MS) and nuclear magnetic resonance (NMR). MS-based technologies, which are generally coupled to chromatographic separations, identify the metabolome components based on their charge-to-mass ratio, have high sensitivity and therefore provide a comprehensive overview of the metabolome. As a drawback, they suffer for reproducibility problems especially in quantitative analyses and require specific approaches for chemically different classes of compounds. In NMR-based metabolomics, only

the signals of all metabolites present in a given sample above a certain concentration threshold are observable, but they are all simultaneously measured in a single monodimensional spectrum. The information about the chemical identity of many of the observed molecules can be retrieved by matching the spectral patterns to those in available databases. At variance with MS, NMR demands little sample preparation and is rapid and quantitative, but has lower sensitivity [3].

CERM, which hosts the daVEB infrastructure, is the Magnetic Resonance Center of the University of Florence and has a long-standing tradition in NMR of biological systems. Since 1994 CERM has been funded by EC to provide transnational European access to its NMR resources; since 2007 it is one of the core labs of the European Research Infrastructure for Integrated Structural Biology (INSTRUCT; www.instruct-fp7.eu) that has been defined by the European Strategy Forum for Research Infrastructure (ESFRI). During the last decade and in collaboration with the Fiorgen Foundation, CERM has established a metabolomics research unit/group and hosts NMR spectrometers that are fully dedicated to metabolomics analyses. Collaborative projects have led to the metabolomics characterization of several pathological states (celiac disease, heart failure, and different types of cancer and other diseases) [4–18]. Very importantly the existence of an individual fingerprint in the urine metabolome of a group of subjects has been identified, which is stable over the time course of years [19,20]. The identification of specific disease signatures in the metabolomics profiles and the ability to detect drifts in the individual metabo-phenotype are expected to have potential applications for early diagnosis [21].

Metabolomics for Biobanks and Biobanks for Metabolomics

There is an obvious link between metabolomics and biobanking activities: Metabolomics profiling and biomarker discovery rely on the comparison of data acquired on large ensembles of samples from different groups of donors. The subsequent analysis allows researchers to assess group-wise differences with statistical approaches that are mainly adapted from earlier emerged omics technologies. Biobanks are obvious providers of samples for large-scale studies and, as such, are a much-needed resource in metabolomics, adding capacity and creating value for the research community.

The biomedical significance of metabolic analyses requires that the metabolome of the sample to be studied represents as closest as possible the original “*in vivo*” metabolome. However, during the entire pre-analytical phase (which covers sample collection, transport, handling, and storage) the concentrations of several metabolites can change due to residual enzymatic activities in the samples and chemical reactions, e.g., induced by exposure to air and/or light. The collaboration between CERM and daVEB has led to the development of metabolomics as an analytical tool to evaluate the effect of different pre-analytical procedures on the sample metabolome [22–24]. Several of these activities were carried out in the context of the FP7 project SPIDIA (<http://www.spidia.eu/>) and have led to development of evidence-based standard operating procedures, now implemented at daVEB, which cover the pre-analytical steps for urine, serum and plasma samples [22,24]. In particular, best practices for urine require that the collected sample is kept refrigerated (4 °C) for a maximum of 2 h and freezing is avoided prior to mild centrifugation and/or filtration to prevent cell disruption upon ice crystal formation [22,24]. Analogously, it has been found that a tight control of temperatures and time during the pre-processing of blood and before freezing of the derived serum and plasma samples is important to preserve the enzymatic degradation and the oxidation of several of the most abundant metabolites [24]. Some critical steps of

tissue preanalytics have been also addressed [20]. For these samples pre- and post-excision ischemia times were demonstrated to be important sources of variability in metabolic profiles. Preservation of sample properties throughout the long-term storage at daVEB is under study; the definition of the maximum life time of a sample in terms of the stability of its small molecule profile has obvious implication for the outcome and reproducibility of metabolomics studies based on samples stored in biobanks and impacts on the budget of biobanks because of the costs associated to sample storage.

For the close links between metabolomics and biobanks, researchers at CERM have promoted the foundation of consortium called EXCEMET (EXpert Center for METabolomics; www.excemet.org) [25], aimed at providing metabolomics services to the previously cited BBMRI-ERIC. As of today, eight parties are members of the consortium: Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine—CIRMMP (Florence, Italy), Medical University of Graz (Graz, Austria), Metanomics Health GmbH (Berlin, Germany), University of Gothenburg (Gothenburg, Sweden), University of Birmingham (Birmingham, UK), Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt—GmbH (München, Germany), Steno Diabetes Center (Gentofte, Denmark), EMBL-EBI (Hinxton, UK).

3. daVEB Collections

The biomaterial available at daVEB includes different types of samples, as summarized in Figure 1a, with a strong focus on serum and urine, as the most common biofluids employed in metabolomics studies. The samples arise from patients with different pathologies (Figure 1b), but an ample collection comes from healthy controls, which are essential to develop models of metabolomic fingerprint of diseases. Donors defined as healthy subjects are individuals that meet the criteria for voluntary blood donation. A more detailed description of the daVEB sample collections is provided as Supplementary Material (see Table A1).

The availability of gender and age-matched controls is an added value for the definition of the diseases signature in metabolomics profiles. The present capacity of the biobank is of about 265,000 vials (0.5, 1 and 2 mL in volume) to be stored at $-80\text{ }^{\circ}\text{C}$ and 195,000 vials (0.5, 1 and 2 mL in volume) to be stored in the vapor of liquid nitrogen, at temperatures below $-170\text{ }^{\circ}\text{C}$. Considering the sample turnover due to sample distribution for research purposes, on average daVEB hosts about 10,000 samples.

The biobank catalogue can be browsed via the web: <https://www.davincieuropeanbiobank.org/en/research/sample-search>. Samples and the associated data, which are project specific, are available for research purposes to external users, according to the published accession criteria [1].

A minimum data set is required for a sample to be accepted in daVEB, which includes the following info: Phenotypic descriptions (donor height, weight, BMI), demographic information (gender, age, ethnicity), individual information (pregnancy, breast-feeding, post-menopausal woman, fasting, smoking habits, alcohol consumption, disease or healthy status, infection state).

Information concerning the sample history is associated to each aliquot, in particular the processing SOP followed at collection, the temperature story (freeze chain) including transport and storage temperatures, and the freeze time, from transport date and duration, entry date in $-80\text{ }^{\circ}\text{C}$ mechanical freezers or tank for storage in the vapor phase of liquid nitrogen, to the registration and exit date.

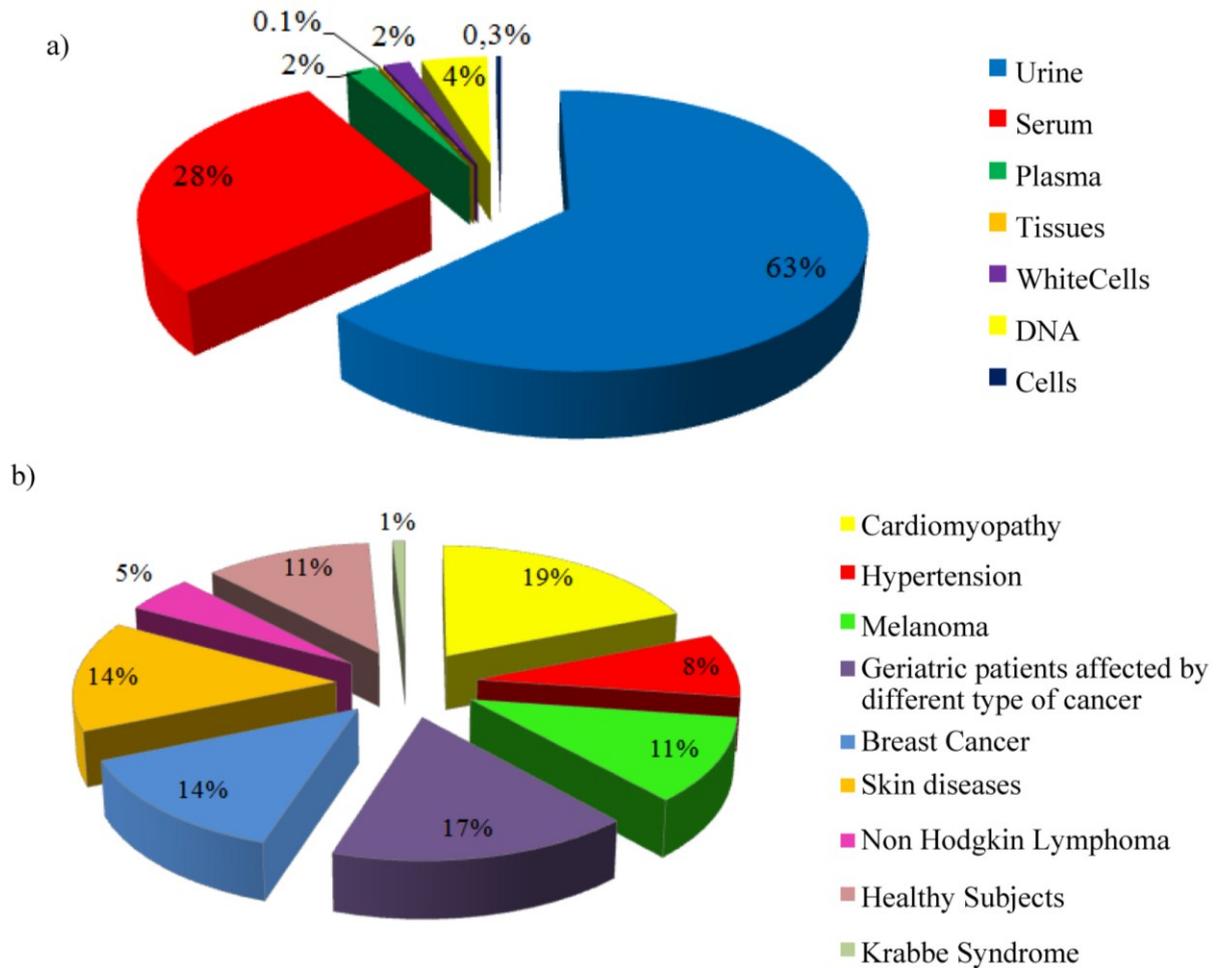


Figure 1. Percentage distribution of da Vinci European BioBank (daVEB) collections by (a) samples type and (b) pathology (healthy subjects included).

4. High Standards in the Data Management System

The technology-driven nature of the da Vinci European BioBank raises the need for a very specific design of its IT infrastructure. In particular, the present design takes into account the following characteristics of the operations of daVEB: (a) the samples stored originate from different collections, which do not share a focus on a common disease nor a common sample type and are collected in different, multiple centers; (b) each collection is associated to a different organization and a different ensemble of clinical data; (c) besides clinical data, samples are associated to other biomedical information, which, in addition to metabolomics data (present for the large majority of samples), can include genetic data, various kinds of images, *etc.* These three characteristics made it necessary for the daVEB biobank to implement a flexible infrastructure permitting the management of different data types for different collections, both for clinical and for other biomedical information. Furthermore, the IT infrastructure allows each collection to be extended at a later moment to include new samples (e.g., for follow-up) or new data types to be associated to the current samples (e.g., the results of analyses not initially available). A further need arising from the above is to implement an approach to search our database that allows users to compare collections with different content types and, sometimes, formats.

A paramount concern when dealing with biobanks is the secure access to the data and the confidentiality of personal data. The daVEB IT infrastructure has addressed this problem by consistently using Open SSL/TSL standard protocols for data exchange and network communications. Furthermore, data are separately stored onto two different servers. All personal information is kept on a server that is not accessible from the public network, but only through a private network and with different credentials than the first server. All other data, which do not raise privacy issues, reside on a separate server reachable via the public network.

For the practical implementation of the daVEB IT infrastructure, we decided to exploit as much as possible open standards and open source software. This should also facilitate the interaction, and eventually the exchange of information, with other biobanks at the European and global level. Our data models are thus available in the form of schemes stored as Open Document Format (ISO Standard) data sheets; as an example, this includes the specific formats and definitions adopted for the information concerning sample history. Regarding data vocabularies and semantics, we designed procedures that allow the daVEB staff to define relationships between different relevant ontologies. We centrally rely on the Open Biological and Biomedical Ontologies made available by the OBO foundry (<http://www.obofoundry.org/>) [26]. However, other ontologies are also acceptable, if they can be accessed as RESTful services [27]. This makes it possible, for example, to use different catalogues of standard disease codes. Links between the corresponding entries of different catalogues can be established in a semi-automated fashion and then subject to manual curation and editing. To define the typical information content associated to a collection and its samples, we took into the core variables defined within the DataSHaPER approach [28]. This should facilitate the harmonization of our data with other studies, also not involving biobanks. On top of this common background, the data schema can be tuned for each sample collection, as detailed in the next paragraph. Finally, we deployed the database management and operating system components using PostgreSQL DBMS and CentOS Linux, respectively. Specific backup and update procedures have also been defined. daVEB is equipped with a fiber channel connecting to a HP MSL6030 Tape Library with two LT04 drives. We implemented a backup schedule that balances lengthier full backups and faster incremental backup strategies, with weekly full backups and daily incremental backups. Moreover, every two months we move to a separate location a full backup of all daVEB data.

More in detail, each collection corresponds to a specific data scheme. An example is shown for a collection of samples from coeliac patients as a table in Figure 2. In the example, to each sample of the collection the system associates SampleCode, Pathology, the number of aliquots available and the number of aliquots already used (e.g., for metabolomics experiments). The field AvailableData contains a link to the collected metabolomics data. Additional such fields can be added to the table, also after its creation, to point to other data types, such as genetic profiling. Additional tables are used in a similar fashion to store the data of donors as well as clinical data.

SampleCode	NTotAliquots	NUsedAliquots	SampleType	Pathology	...	AvailableData
COE00010SE001			Serum	coeliac disease	...	COE00010SE001_filepath
COE00001UR001			Urine	coeliac disease	...	COE00001UR001_filepath
COE00002SE001			Serum	coeliac disease	...	COE00002SE001_filepath
...						

Figure 2. Scheme of the data table for a collection of samples from coeliac patients. The SampleCode field uniquely identifies each sample of the collection.

The meaning of the fields in a table such as that of Figure 2 is defined in higher-level metadata table, as shown in Figure 3. Such table contains the definition of the metadata associated to each collection of the daVEB biobank. This is done by defining for each field of the data table of the collection the name of the field, its reference to a specific ontology, and the type of data (numeric, string, link, etc.) that can be associated to each field (see Figure 3). Each record of the metadata table corresponds to a specific field (defined by the combination of TableRef and FieldRef fields). The OntoRef field indicates the ontology used for that particular field of the data table, which can be an internally defined ontology or a public reference ontology. The reference used for the field is indicated by the DataRef field, whereas the allowed data type is defined by the daVEBDescr field. Each data type, as mentioned, can be a single instance field, e.g., “Coeliac Disease”, or a type set. If the ontology used is an external one, then the DataRef field can point directly to external RESTful ontology, as shown in the example. This allows the unequivocal definition of the term “Coeliac Disease”. However, it is possible to use multiple ontologies, if necessary, by providing multiple records for each term, with pointers to the appropriate external vocabularies.

TableRef	FieldRef	OntoRef	DataRef	daVEBDescr
COESAMPLETAB	SAMPLECODE	DAVEBCODES	SAMPLECODERef	String Object
COESAMPLETAB	NTOTALIQUOTS	DAVEBTYPES	INTEGERRef	Integer Object
...				
COESAMPLETAB	PATHOLOGY	DOID (Human Diseases)	DOID:10608	“Coeliac disease”
COESAMPLETAB	AVAILABLEDATA	DAVEBRESOURCES	FILEPATHRef	String Object
...				

Term Name: (Include obsolete terms) Term ID:

Coeliac disease

Additional Information:

definition	A food allergy that is caused by a reaction located in small intestine to gliadin, a prolamin (gluten protein) found in wheat, and similar proteins found in the crops of the tribe Triticeae. The disease is associated with HLA-DQ gene. It has_symptom abdominal pain, has_symptom constipation, has_symptom diarrhea, has_symptom nausea and vomiting, and has_symptom loss of appetite.
preferred name	celiac disease
exact synonym	celiac sprue [CSP2005:1248-3893]
exact	irritation; steatorrhea [MTHICD9: 2006:579.01]

Figure 3. Example of first-level daVEB metadata table.

At the top of the metadata tables of the daVEB stack there is the Ontologies table (Figure 4), where every single ontology and standard data set is referenced through an appropriate HTTP path to allow remote procedure calls exploiting RESTful services.

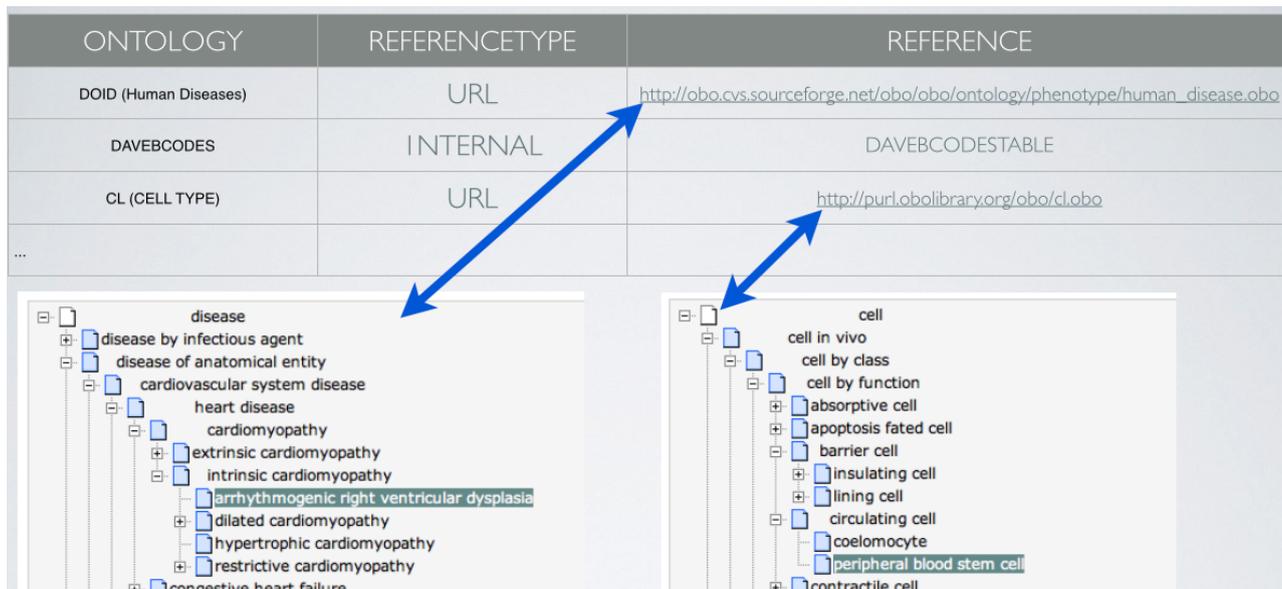


Figure 4. The daVEB Ontologies table.

Let us now consider the case of a user who wants to compare the data contained in two different collections. Such projects will have different data tables and descriptors. In a more traditional setup than that of the daVEB biobank, the corresponding data would most likely be collected into different databases adopting different data schemes. This however would make it difficult and inefficient to merge queries over the two collections of interest in order to discover, for example, similarities/differentiations among the life habits of donors. The daVEB approach allows us to combine queries for each single project into a single query using the standard SQL language for queries. Such queries exploit the relations among metadata definitions contained in the local tables or accessible via public RESTful services (see Figure 4).

Therefore, we can easily define a query relating data among two different projects, e.g., the COE and MET projects:

find all

(A) diet habit elements for healthy subjects in MET project

that are not in

(B) diet habit elements for pathologic subjects in COE project

We can translate such a question into a nested SQL query using the data contained in metadata tables to capture common definitions for diet habit in the MET and COE projects, and use such information to determine the result set, defined as the difference A-B. The entire data extraction procedure is automated and entirely SQL driven.

As far as the development of standards for the inclusion of metabolomics data in the biobank database, the involvement of some of us in the FP7 project COSMOS has been particularly valuable. COSMOS is a collaborative project funded by the European Commission within the 7th Framework programme

(<http://www.cosmos-fp7.eu/>). The specific aim of COSMOS is to develop open standards for the exchange of metabolomics data and the associated metadata, together with a variety of tools to support capturing and management of such information. These tools interface to the software used for data acquisition, to commonly used programs for the analysis of metabolomics data, and to relevant databases, such as Metabolights (<http://www.ebi.ac.uk/metabolights/>) [29]. COSMOS has been developing its standards in collaboration with the major instrumentation vendors as well as publishers in the field, and relevant international organizations. As mentioned, a key specificity of the daVEB biobank is the availability of NMR-based metabolomics information for the large majority of its samples, especially urine and blood. The interaction with the COSMOS project is thus instrumental to make such information as widely reusable as possible. In this respect, the development of nmrML (<http://nmrml.org>) as an open mark-up language specifically designed for NMR data resulting from metabolomics applications has been a crucial achievement. Notably, it is not needed to store the experimental data in the nmrML format, as they can be directly imported at a later stage. On the other hand, the ontologies used within the daVEB data warehouse are compatible with the ISA software suite that is central to metadata management in COSMOS. COSMOS has been developing a controlled vocabulary for NMR experiments, called nmrCV, adopting the OWL (Web Ontology Language, developed by the W3C consortium) syntax. The OBO ontology used within the IT infrastructure of daVEB can be readily converted to the OWL format.

Beyond the mere distribution of the metabolomics profiles associated with the various samples, the daVEB biobank provides support for the reuse of the derived biomedical and clinical information. For example, COSMOS provides a computational infrastructure to identify and quantitate specific metabolites within the entire profile. This information can be derived from the metabolomics data in the daVEB biobank through a variety of analytic approaches and its significance is defined on a statistical basis [30]. However, this requires the involvement of an expert skilled in metabolomics to perform the analysis. In the future it may become feasible for daVEB users to directly access the derived information, through the COSMOS framework, thus avoiding the need for repeating the computational analysis, unless desired/needed.

5. Conclusions

The rapid development of high-throughput technologies and computational frameworks enables the examination of biological systems in unprecedented detail, paving the way to significant advances in personalized and precision medicine. By providing tissues, cells, fluids, and raw data, biobanks have become the drivers of the translational research cycle, since they are the custodians of the biospecimens and related anonymized data. A deep knowledge of the requirements inherent to a specific omic analytical approach allows biobanks to foster the development of this discipline. The close collaboration between biobanks and analytical platforms becomes essential to create a new generation of biobanks acting as drivers of the development of new methodologies.

Acknowledgments

We acknowledge the generous financial support of Fondazione Cassa di Risparmio di Firenze, which provided the initial investment for the establishment of daVEB (Contract No. 2007.1324), its expansion

(Contracts Nos. 2008.2026 and 2009.1165) and the funding of specific research projects for the determination of biobanked sample quality (contracts Nos. 2013.0494 and 2011.1073).

Further, we would like to thank the Regione Toscana, Grant No. 2010_FI_18_FIPRO2 for the project “Rete di Infrastrutture industriale per l’incubazione e per servizi avanzati alle imprese innovative e locali (RISE)—Network for productive Infrastructures”.

Author Contributions

A.R. and P.T. drafted the manuscript and conceptually designed the manuscript.

C.L. critically revised the manuscript.

D.C. and G.M. contributed data and prepared the graphical material and supporting information.

G.M. took care of the editorial aspects.

Appendix

Table A1. The different kind of biosamples available for each disease, as absolute numbers.

Disease	Donor number	Sample type	Sample number	Available aliquots
Cardiomyopathy	291	Urine	776	2863
		Serum	810	2492
Hypertension	126	Serum	126	378
Melanoma	156	Urine	153	892
	169	Serum	174	1137
	166	Plasma	170	1132
	154	White Cells	156	454
	4	Tissue	5	12
Geriatric patients affected by different type of cancer	257	Urine	514	941
		Serum	257	263
Breast Cancer	209	Urine	1520	2150
		Serum	735	412
Skin diseases	219	Serum	219	254
	2	Plasma	2	2
	21	Cells	27	27
	3	Tissue	3	3
Non Hodgkin Lymphoma	73	DNA	381	402
Healthy Subjects	175	Urine	2476	9604
		Serum	149	170
Krabbe Syndrome	13	Urine	10	79
		Serum	11	45

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Marcon, G.; Nincheri, P. The Multispecialistic da Vinci European BioBank. *Open J. Bioresour.* **2014**, *1*, e6.
2. Wishart, D.S.; Jewison, T.; Guo, A.C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; *et al.* HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Res.* **2013**, *41*, D801–D807.
3. Veenstra, T.D. Metabolomics: The final frontier? *Genome Med.* **2012**, *4*, doi:10.1186/gm339.
4. Calabrò, A.; Gralka, E.; Luchinat, C.; Saccenti, E.; Tenori, L. A Metabolomic Perspective on Coeliac Disease. *Autoimmune Dis.* **2014**, *2014*, doi:10.1155/2014/756138.
5. Bernini, P.; Bertini, I.; Calabrò, A.; la Marca, G.; Lami, G.; Luchinat, C.; Renzi, D.; Tenori, L. Are patients with potential celiac disease really potential? The answer of metabonomics. *J. Proteome Res.* **2011**, *10*, 714–721.
6. Bertini, I.; Calabrò, A.; de Carli, V.; Luchinat, C.; Nepi, S.; Porfirio, B.; Renzi, D.; Saccenti, E.; Tenori, L. The metabonomic signature of celiac disease. *J. Proteome Res.* **2009**, *8*, 170–177.
7. Padeletti, L.; Modesti, P.A.; Cartei, S.; Checchi, L.; Ricciardi, G.; Pieragnolia, P.; Sacchi, S.; Padeletti, M.; Alterini, B.; Pantaleo, P.; *et al.* Metabolomic does not predict response to cardiac resynchronization therapy in patients with heart failure. *J. Cardiovasc. Med. (Hagerstown)* **2014**, *15*, 295–300.
8. Tenori, L.; Hu, X.; Pantaleo, P.; Alterini, B.; Castelli, G.; Olivotto, I.; Bertini, I.; Luchinat, C.; Gensini, G.F. Metabolomic fingerprint of heart failure in humans: A nuclear magnetic resonance spectroscopy analysis. *Int. J. Cardiol.* **2013**, *168*, doi:10.1016/j.ijcard.2013.08.042.
9. Bernini, P.; Bertini, I.; Luchinat, C.; Tenori, L.; Tognaccini, A. The cardiovascular risk of healthy individuals studied by NMR metabonomics of plasma samples. *J. Proteome Res.* **2011**, *10*, 4983–4992.
10. Tenori, L.; Oakman, C.; Morris, P.G.; Gralka, E.; Turner, N.; Cappadona, S.; Fornier, M.; Hudis, C.; Norton, L.; Luchinat, C.; *et al.* Serum metabolomic profiles evaluated after surgery may identify patients with oestrogen receptor negative early breast cancer at increased risk of disease recurrence. Results from a retrospective study. *Mol. Oncol.* **2015**, *9*, 128–139.
11. Oakman, C.; Tenori, L.; Cappadona, S.; Luchinat, C.; Bertini, I.; di Leo, A. Targeting metabolomics in breast cancer. *Curr. Breast Cancer Rep.* **2012**, *4*, 249–256.
12. Tenori, L.; Oakman, C.; Claudino, W.M.; Bernini, P.; Cappadona, S.; Nepi, S.; Biganzoli, L.; Arbushites, M.C.; Luchinat, C.; Bertini, I.; *et al.* Exploration of serum metabolomic profiles and outcomes in women with metastatic breast cancer: A pilot study. *Mol. Oncol.* **2012**, *6*, 437–444.
13. Oakman, C.; Tenori, L.; Claudino, W.M.; Cappadona, S.; Nepi, S.; Battaglia, A.; Bernini, P.; Zafarana, E.; Saccenti, E.; Fornier, M.; *et al.* Identification of a serum-detectable metabolomic fingerprint potentially correlated with the presence of micrometastatic disease in early breast cancer patients at varying risks of disease relapse by traditional prognostic methods. *Ann. Oncol.* **2011**, *22*, 1295–1301.
14. Claudino, W.M.; Quattrone, A.; Biganzoli, L.; Pestrin, M.; Bertini, I.; di Leo, A. Metabolomics: Available results, current research projects in breast cancer, and future applications. *J. Clin. Oncol.* **2007**, *25*, 2840–2846.

15. Bertini, I.; Cacciatore, S.; Jensen, B.V.; Schou, J.V.; Johansen, J.S.; Kruhøffer, M.; Luchinat, C.; Nielsen, D.L.; Turano, P. Metabolomic NMR fingerprinting to identify and predict survival of patients with metastatic colorectal cancer. *Cancer Res.* **2012**, *72*, 356–364.
16. Dani, C.; Bresci, C.; Berti, E.; Ottanelli, S.; Mello, G.; Mecacci, F.; Breschi, R.; Hu, X.; Tenori, L.; Luchinat, C. Metabolomic profile of term infants of gestational diabetic mothers. *J. Matern. Fetal Neonatal Med.* **2014**, *27*, 537–542.
17. Bertini, I.; Luchinat, C.; Miniati, M.; Monti, S.; Tenori, L. Phenotyping COPD by ¹H-NMR metabolomics of exhaled breath condensate. *Metabolomics* **2014**, *10*, 302–311.
18. Aimetti, M.; Cacciatore, S.; Graziano, A.; Tenori, L. Metabonomic analysis of saliva reveals generalized chronic periodontitis signature. *Metabolomics* **2012**, *8*, 465–474.
19. Assfalg, M.; Bertini, I.; Colangiuli, D.; Luchinat, C.; Schäfer, H.; Schütz, B.; Spraul, M. Evidence of different metabolic phenotypes in humans. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1420–1424.
20. Bernini, P.; Bertini, I.; Luchinat, C.; Nepi, S.; Saccenti, E.; Schäfer, H.; Schütz, B.; Spraul, M.; Tenori, L. Individual human phenotypes in metabolic space and time. *J. Proteome Res.* **2009**, *8*, 4264–4271.
21. Bertini, I.; Luchinat, C.; Tenori, L. Metabolomics for the future of personalized medicine through information and communication technologies. *Pers. Med.* **2012**, *9*, 133–136.
22. Emwas, A.-H.; Luchinat, C.; Turano, P.; Tenori, L.; Roy, R.; Salek, R.M.; Ryan, D.; Merzaban, J.S.; Kaddurah-Daouk, R.; Zeri, A.C.; *et al.* Standardizing the experimental conditions for using urine in NMR-based metabolomic studies with a particular focus on diagnostic studies: A review. *Metabolomics* **2014**, doi:10.1007/s11306-014-0746-7.
23. Cacciatore, S.; Hu, X.; Viertler, C.; Kap, M.; Bernhardt, G.A.; Mischinger, H.J.; Riegman, P.; Zatloukal, K.; Luchinat, C.; Turano, P. Effects of intra- and post-operative ischemia on the metabolic profile of clinical liver tissue specimens monitored by NMR. *J. Proteome Res.* **2013**, *12*, 5723–5729.
24. Bernini, P.; Bertini, I.; Luchinat, C.; Nincheri, P.; Staderini, S.; Turano, P. Standard Operating Procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. *J. Biomol. NMR* **2011**, *49*, 231–243.
25. Van Ommen, G.J.; Törnwall, O.; Bréchet, C.; Dagher, G.; Galli, J.; Hveem, K.; Landegren, U.; Luchinat, C.; Metspalu, A.; Nilsson, C.; *et al.* BBMRI-ERIC as a resource for pharmaceutical and life science industries: The development of biobank-based Expert Centres. *Eur. J. Hum. Genet.* **2014**, doi:10.1038/ejhg.2014.235.
26. Smith, B.; Ashburner, M.; Rosse, C.; Bard, J.; Bug, W.; Ceusters, W.; Goldberg, L.J.; Eilbeck, K.; Ireland, A.; Mungall, C.J.; *et al.* The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* **2007**, *25*, 1251–1255.
27. Richardson, L.; Ruby, S. *RESTful Web Services—Web Services for the Real World*, 1st ed.; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2007.
28. Fortier, I.; Burton, P.R.; Robson, P.J.; Ferretti, V.; Little, J.; L'Heureux, F.; Deschênes, M.; Knoppers, B.M.; Doiron, D.; Keers, J.C.; *et al.* Quality, quantity and harmony: The DataSHaPER approach to integrating data across bioclinical studies. *Int. J. Epidemiol.* **2010**, *39*, 1383–1393.

29. Haug, K.; Salek, R.M.; Conesa, P.; Hastings, J.; de Matos, P.; Rijnbeek, M.; Mahendraker, T.; Williams, M.; Neumann, S.; Rocca-Serra P.; *et al.* MetaboLights—An open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.* **2013**, *41*, D781–D786.
30. Cacciatore, S.; Luchinat, C.; Tenori, L. Knowledge discovery by accuracy maximization. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 5117–5122.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).