Title: Towards more transparent and reproducible omics studies through a common metadata checklist and data publications

Running Title:
Metadata checklist for omics studies and data publications

Keywords:
Multi-disciplinary, multi-omics, metadata, checklist, guidelines, transparency, reproducibility, integrated analysis, data, reuse

Authors:

Eugene Kolker*
Bioinformatics and High-Throughput Analysis Laboratory,
Seattle Children's Research Institute,
Predictive Analytics, Seattle Children's
1900 Ninth Avenue
Seattle, WA 98101
Email: eugene.kolker@seattlechildrens.org
Phone: 206-884-7170

Vural Özdemir
Office of the President, Gaziantep University,
International Affairs and Global Development Strategy,
and Faculty of Communications
Üniversite Bulvarı P.K. 27310 Şehitkamil
Gaziantep, Kilis Yolu, Turkey
Email: vural.ozdemir@gantep.edu.tr
Phone: 90-545-550-88-80

Lennart Martens
Department of Medical Protein Research,
VIB, Ghent, Belgium
Department of Biochemistry,
Ghent University, Ghent, Belgium
Albert Baertsoenkaai 3
9000 GENT
Email: lennart.martens@vib-ugent.be
Phone: 32-9-264-93-58

William Hancock
Barnett Institute, Department of Chemistry
Northeastern University

341 Mugar Building
Boston MA 02155
Email: wi.hancock@neu.edu
Phone: 617-373-4881

Gordon Anderson
Fundamental & Computational Sciences Directorate
Pacific Northwest National Laboratory
902 Battelle Boulevard
P.O. Box 999, MSIN K8-98
Richland, WA 99352
Email: Gordon@pnl.gov
Phone: 509-371-6582

Nathaniel Anderson
Bioinformatics and High-Throughput Analysis Laboratory,
Seattle Children's Research Institute
1900 Ninth Avenue
Seattle, WA 98101
Email: nate.anderson@seattlechildrens.org
Phone: 206-884-1230

Şükrü Aynacioglu
Department of Pharmacology
Gaziantep University
University Boulevard, 27310 Sehitkamil
Gaziantep, Turkey
Email: aynacioglu@gantep.edu.tr
Phone: 90-342-360-1023

Ancha Baranova
School of Systems Biology
George Mason University
10900 University Blvd, MSN 5B3
Manassas, VA 20110
Email: abaranov@gmu.edu
Phone: 703-993-4293

Shawn R. Campagna
Department of Chemistry
University of Tennessee, Knoxville
618 Buehler Hall
Knoxville, TN 37996-1600
Email: campagna@utk.edu
Phone: 865-974-7337

Rui Chen
Department of Genetics
Stanford University
300 Pasteur Drive, Alway M-308
Stanford, CA 94305-5120
Email: ruichens@stanford.edu
Phone: 650-723-3277

John Choiniere
Bioinformatics and High-Throughput Analysis Laboratory
Seattle Children's Research Institute
1900 Ninth Avenue
Seattle, WA 98101
Email: john.choiniere@seattlechildrens.org
Phone: 206-884-1251

Stephen P. Dearth
Department of Chemistry
University of Tennessee, Knoxville
1420 Circle Dr.
Knoxville, TN 37996
sdearth@ion.chem.utk.edu
Phone: 865-974-3141

Wu-Chun Feng
SyNeRGy Laboratory
Departments of Computer Science & Electrical and Computer Engineering
Virginia Tech
2202 Kraft Dr
Blacksburg, VA 24060
Email: feng@cs.vt.edu
Phone: 540-231-1192

Lynnette Ferguson
Auckland Cancer Society Research Centre
Department of Nutrition
University of Auckland
Private Bag 92019, Auckland Mail Centre
Auckland 1142
New Zealand
Email: l.ferguson@auckland.ac.nz
Phone: 64-9-373-7599, ext. 86372

Geoffrey Fox
School of Informatics and Computing
Indiana University

815 E 10th Street, Room 105
Bloomington, IN 47408
Email: gcf@indiana.edu
Phone: 812-856-7977

Dmitrij Frishman
Technische Universität München,
Wissenshaftzentrum Weihenstephan ,A
m Forum 1, 85354 Freising, Germany
Email: d.frishman@wzw.tum.de
Phone: 49-8161-712134

Robert Grossman
Institute for Genomics and Systems Biology
and Department of Medicine,
University of Chicago
900 East 57th Street, Room 10142
Chicago, IL 60637
Email: robert.grossman@uchicago.edu
Phone: 773-834-4669

Allison Heath
Institute for Genomics and Systems Biology,
University of Chicago
Knapp Center for Biomedical Discovery
900 East 57th Street
Chicago, IL 60637
Email: aheath@uchicago.edu
Phone: 773 573-9560

Roger Higdon
Bioinformatics and High-Throughput Analysis Laboratory,
Seattle Children's Research Institute
Predictive Analytics, Seattle Children's
1900 Ninth Avenue
Seattle, WA 98101
Email: roger.higdon@seattlechildrens.org
Phone: 206-884-7172

Mara H. Hutz
Departamento de Genética,
Instituto de Biociências,
Federal University of Rio Grande do Sul,
Caixa Postal 15053, 91501-970
Porto Alegre, RS, Brazil.
E-mail: mara.hutz@ufrgs.br

Imre Janko
High-Throughput Analysis Core,
Seattle Children's Research Institute
1900 Ninth Avenue
Seattle, WA 98101
Email: imre.janko@seattlechildrens.org
Phone: 206-884-1064

Lihua Jiang
Department of Genetics
Stanford University
300 Pasteur Drive, M-344A
Stanford, California 94305
Email: lihuaj@stanford.edu
Phone: 650-723-9914

Sanjay Joshi
Life Sciences
EMC^2
Email: sanjay.joshi@emc.com
Phone: 425-429-4727

Alexander Kel
GeneXplain GmbH
Am Exer 10 b
D-38302 Wolfenbüttel
Germany
Email: alexander.kel@genexplain.com
Phone: +49-05331-992200-13

Joseph W. Kemnitz
Department of Cell and Regenerative Biology
Wisconsin National Primate Research Center
University of Wisconsin-Madison
1220 Capitol Court
Madison, WI 53715-1299
Email: kemnitz@primate.wisc.edu
Phone: 608-263-3588

Isaac S. Kohane
Pediatrics and Health Sciences and Technology
Children's Hospital and Harvard Medical School
HMS Center for Biomedical Informatics
Countway Library of Medicine
10 Shattuck Street

Boston, MA 02115
Email: isaac_kohane@harvard.edu
Phone: 617-432-2144

Natali Kolker
High-Throughput Analysis Core,
Seattle Children's Research Institute
Predictive Analytics, Seattle Children's
1900 Ninth Avenue
Seattle, WA 98101
Email: natali.kolker@seattlechildrens.org
Phone: 206-884-7171

Doron Lancet
Crown Human Genome Center
Department of Molecular Genetics
Weizmann Institute of Science
Rehovot 76100, Israel
Email: doron.lancet@weizmann.ac.il
Phone: 972-8-934-3683

Elaine Lee
High-throughput Analysis Core,
Seattle Children's Research Institute
1900 Ninth Avenue
Seattle, WA 98101
Email: Elaine.lee@seattlechildrens.org
Phone: 206-884-1250

Weizhong Li
Center for Research in Biological Systems
University of California, San Diego
9500 Gilman Drive MC 0446
Atkinson Hall, Room 3113
La Jolla, CA 92093-0446
Email: liwz@sdsc.edu
Phone: 858-534 4143

Andrey Lisitsa
Russian Human Proteome Organization (RHUPO)
Institute of Biomedical Chemistry
Moscow, 199121 Russia
Pogodinskaya str., 10
Email: Lisitsa063@gmail.com
Phone: 7-499-246-3731

Adrián Llerena
Clinical Research Center
Extremadura University Hospital and Medical School
Hospital Infanta Cristina Av de Elvas s/n
06080 Badajoz, Spain
Email: allerena@unex.es
Phone: +34924218040

Courtney MacNealy-Koch
DELSA Global,
Seattle Children's Research Institute
1900 Ninth Avenue
Seattle, WA 98101
Email: courtney.macnealykoch@seattlechildrens.org
Phone: 206-884-1171

Jean-Claude Marshall
Center for Translational Research
Catholic Health Initiatives
7601 Osler Drive,
Towson, MD 21204
Email: Jean-ClaudeMarshall@catholichealth.net
Phone: 410.427.2587

Paola Masuzzo
Department of Medical Protein Research
VIB, Ghent, Belgium
Department of Biochemistry
Ghent University, Ghent, Belgium
Albert Baertsoenkaai 3
9000 GENT
Email: paola.masuzzo@vib-ugent.be
Phone: 32-9-264-93-33

Amanda L. May
Department of Chemistry
UT Knoxville
1420 Circle Dr.
Knoxville, TN 37996-1600
Email: amay@ion.chem.utk.edu
Phone: 865-974-3141

George Mias
Department of Genetics
Stanford University
300 Pasteur Drive, Alway Building M308

Stanford, CA 94305-5120
Email: george.mias@stanford.edu
Phone: 650-723-3277

Matthew Monroe
Biological Sciences Division
Pacific Northwest National Laboratory
902 Batelle Boulevard
P.O. Box 999 MSIN K8-98
Richland, WA 99352
Email: matthew.monroe@pnnl.gov
Phone: 509-371-6580

Elizabeth Montague
Bioinformatics and High-Throughput Analysis Lab,
Seattle Children's Research Institute
Predictive Analytics, Seattle Children's
1900 Ninth Avenue
Seattle, WA 98101
Email: elizabeth.montague@seattlechildrens.org
Phone: 206-884-1137

Sean Mooney
The Buck Institute for Research on Aging
8001 Redwood Blvd.
Novato, CA 94945
Email: smooney@buckinstitute.org
Phone: 415-209-2038

Alexey Nesvizhskii
Department of Pathology,
Department of Computational Medicine and Bioinformatics
University of Michigan
4237 Med Sci I, 1301 Catherine Road
Ann Arbor, MI 48109-0602
Email: nesvi@med.umich.edu
Phone: 734)-764-3516

Santosh Noronha
Department of Chemical Engineering
Indian Institute of Technology Bombay
Powai, Mumbai
400076, INDIA
Email: noronha@iitb.ac.in
Phone: 91-22-2576 7238

Gilbert Omenn
Center for Computational Medicine and Bioinformatics
Departments of Molecular Medicine & Genetics and Human Genetics
Department of Computational Medicine and Bioinformatics
School of Public Health
100 Washtenaw Avenue
2017 Palmer Commons
Ann Arbor, MI 48109-2218
Email: gomenn@umich.edu
Phone: 734-763-7583

Harsha K. Rajasimha
Jeeva Informatics Solutions LLC
7706 Majestic Way
Derwood, MD USA
Email: harsha@jeevadx.com
Phone: 540-239-0465

Preveen Ramamoorthy
Advanced Diagnostic Laboratories,
Department of Medicine National Jewish Health
1400 Jackson St
Denver, CO 80206
Email: RamamoorthyP@NJHealth.org
Phone: 303-398-1501

Jerry Sheehan
California Institute for Telecommunications and Information Technology
University of California, San Diego
9500 Gilman Drive #0436
La Jolla, CA 92093-0436
Email: jsheehan@ucsd.edu
Phone: 858-534-1723

Larry Smarr
California Institute for Telecommunications and Information Technology
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0436
Email: lsmarr@ucsd.edu
Phone: 858-822-4284

Charles V. Smith
Center for Developmental Therapeutics
Seattle Children's Research Institute
1900 Ninth Avenue

Seattle, WA 98101
Email: charlesskip.smith@seattlechildrens.org
Phone: 206-884-7453

Todd Smith
Digital World Biology
2442 NW Market St. PMB 160
Seattle WA, 98107
Email: todd@digitalworldbiology.com
Phone: 206-375-6843

Michael Snyder
Department of Genetics
Stanford Center for Genomics and Personalized Medicine
Stanford University
300 Pasteur Drive, M-344A
Stanford, California 94305
Email: mpsnyder@Stanford.edu
Phone: 650-736-8099

Srikanth Rapole
Proteomics Laboratory
National Centre for Cell Science
NCCS Complex, Pune University Campus
Ganeshkhind, Maharashtra
Pune- 411007, India
Email: rsrikanth@nccs.res.in
Phone: +91-20-25708075

Sanjeeva Srivastava
Proteomics Laboratory
Indian Institute of Technology Bombay
Powai, Mumbai
400 076, India
Email: sanjeeva@iitb.ac.in
Phone: 022-2576-7779

Larissa Stanberry
Bioinformatics and High-Throughput Analysis Laboratory,
Seattle Children's Research Institute
Predictive Analytics, Seattle Children's
1900 Ninth Avenue
Seattle, WA 98101
Email: larissa.stanberry@seattlechildrens.org
Phone: 206-884-1059

Elizabeth Stewart
Bioinformatics and High-Throughput Analysis Laboratory
Seattle Children's Research Institute
1900 Ninth Avenue
Seattle, WA 98101
Email: elizabeth.stewart@seattlechildrens.org
Phone: 206-884-7176

Stefano Toppo
Department of Molecular Medicine
University of Padova
Via U. Bassi 58/b
I-35131 Padova, Italy
Email: stefano.toppo@unipd.it
Phone: 39-049-8276958

Peter Uetz
J Craig Venter Institute (JCVI)
9704 Medical Center Drive
Rockville, MD 20850
Email: uetz@jcvi.org
Phone: 301-795-7589

Kenneth Verheggen
Department of Medical Protein Research
VIB, Ghent, Belgium
Department of Biochemistry
Ghent University, Ghent, Belgium
Albert Baertsoenkaai 3
9000 GENT
Email: kenneth.verheggen@ugent.be
Phone: 32-9-264-93-58

Brynn H. Voy
Department of Animal Science
UT Institute of Agriculture
2506 River Drive
Knoxville, Tennessee 37996
Email: bhvoy@utk.edu
Phone: 865-974-4729

Louise Warnich
Department of Genetics
Faculty of AgriSciences
University of Stellenbosch
Private Bag X1,

Matieland, 7602,
Stellenbosch, South Africa
E-mail: lw@sun.ac.za
Phone: 27-21-808-5888

Steven W. Wilhelm
Department of Microbiology
University of Tennessee, Knoxville
M409 Walters Life Sciences
Knoxville, Tennessee 37996-0845
Email: Wilhelm@utk.edu
Phone: 865-974-0665

Gregory Yandl
Bioinformatics and High-Throughput Analysis Laboratory,
Seattle Children's Research Institute
Predictive Analytics, Seattle Children's
1900 Ninth Avenue
Seattle, WA 98101
Email: gregory.yandl@seattlechildrens.org
Phone: 206-884-3009

*Signifies corresponding author

Abstract

Biological processes are fundamentally driven by complex interactions between biomolecules. Integrated high-throughput omics studies enable multifaceted views of cells, organisms, or their communities. With the advent of new post-genomics technologies omics studies are becoming increasingly prevalent yet the full impact of these studies can only be realized through data harmonization, sharing, meta-analysis, and integrated research,. These three essential steps require consistent generation, capture, and distribution of the metadata. To ensure transparency, facilitate data harmonization, and maximize reproducibility and usability of life sciences studies, we propose a simple common omics metadata checklist. The proposed checklist is built on the rich ontologies and standards already in use by the life sciences community. The checklist will serve as a common denominator to guide experimental design, capture important parameters, and be used as a standard format for stand-alone data publications. This omics metadata checklist and data publications will create efficient linkages between omics data and knowledge-based life sciences innovation and importantly, allow for appropriate attribution to data generators and infrastructure science builders in the post-genomics era. We ask that the life sciences community test the proposed omics metadata checklist and data publications and provide feedback for their use and improvement.

A COMMON OMICS METADATA CHECKLIST PROPOSAL

Modern life science technologies enable rapid and efficient acquisition of omics data. These data comprehensively measure multi-layered molecular networks and provide a snapshot of biological processes in a cell, organism, or their communities. Collected on the same sample at the same time, omics data provide information on the functioning of biomolecules and their interactions. Omics studies are essential for the systemic investigation of biological systems--an endeavor that is crucial to improve our ability to manage and cure diseases, identify drug targets, understand regulatory cascades, and predict ecosystem responses to environmental changes.

Through the pioneering efforts of Drs. Smarr and Snyder (Smarr 2012, Bowden 2012, Chen 2012, Mias 2013), two powerful multi-omics human datasets were recently made available. Smarr's dataset includes a wide variety of molecular measures and clinical parameters meticulously collected and cataloged for years, while Snyder's integrative personal multi-omics study presents his personal genomics, transcriptomics, proteomics, metabolomics, and autoantibody profiles collected over a 14-month period. Both studies yielded unique physiological insights not previously possible, including early indications of vulnerabilities to specific diseases.

In the near future these kinds of personal omics studies will become routine and will inevitably result in vast and diverse volumes of omics data. Therefore, the scientific community must commit to a common format for publishing the design and analysis of these studies that will ensure the compatibility, reproducibility, and reuse of the resulting data.

The use, integration, and reuse of data require accurate and comprehensive capture of the associated metadata including details describing experimental design, sample acquisition and preparation, instrument protocols, and processing steps. The data and metadata must be captured together in a rigorous and consistent manner to allow the integration of data across omics experiments. The use of ontologies, naming conventions and standards can increase the compatibility and usability of these diverse data. Fortunately, life sciences data have certain core similarities. However, combined with these similarities come the different nuances among various technology platforms such as transcriptomics, proteomics and metabolomics, as well as application contexts such as neuroscience and hematology. The differences are compounded by the multiplicity of standards within a field--transcriptomics alone has at least 15 standards potentially applicable to the data (Tennenbaum 2013, Field 2009). Such complexities not only make reproducible, integrative, accurate, and comprehensive capture of data and metadata an intricate challenge that must be overcome but also place an excessive burden on researchers trying to convey metadata (Tennenbaum 2013, Editorial 2011).

Pioneering attempts in this area were made in 2007, when the “Minimum Information about a Biomedical or Biological Investigation" project brought many of these efforts for the life sciences together into an umbrella organization: MIBBI (Taylor 2008, Kettner 2010). In MIBBI, each set of guidelines is developed by a working group concentrated in a specific field, (for example, fMRI, or QTL and association studies). Through this approach, MIBBI aspires to capture all essential metadata and data that are necessary to replicate any given experiment within a field. Also the framework known as Minimal Information about any Sequence (MIxS) expands the breadth of information available by integrating the individual genomics checklists

developed by the Genomics Standards Consortium with environmental information (Yilmaz 2011). In addition, the NIH's National Center for Biotechnology Information developed a format for cataloging information about samples enabling further metadata availability (Barrett 2012). While these frameworks are critical to the reuse of data, they do not fully take into account the interlocking aspects needed for harmonization of diverse omics data types.

Recently, the *Nature* Publishing Group implemented a publication checklist that provides another example of an approach to improve the transparency and reproducibility of life sciences publications (Editorial 2013, Reporting Checklist For Life Sciences Articles 2013). The checklist requires the researcher and/or corresponding author to enter specific information on experimental design, statistical analysis, and reagents. This checklist is endorsed by the Data-Enabled Life Sciences Alliance (DELSA Global, (Kolker 2013)).

The unveiling of the *Nature* publication initiative brought into focus the need for a complementary omics checklist that allows the capture and publication of critical metadata associated with omics data sets. To this end, life sciences researchers from DELSA Global (Data-Enabled Life Sciences Alliance 2013a, Kolker 2012a, Kolker 2012c, Kolker 2012b, Stewart 2013) propose a single common omics metadata checklist as described below. By integrating DELSA researchers' collective experiences with omics guidelines and publication requirements, one simplified, yet informative and flexible checklist was created to capture the essential aspects of omics studies (Data-Enabled Life Sciences Alliance 2013b).

Publication of a completed checklist will serve to inform the life sciences community of the details needed to properly utilize the given data set. This type of "resource publication" has long been done by *Nucleic Acid Research* in its annual Database issue. Nanopubs and Micropubs are two newer publication avenues that could serve to quickly and accurately share information (Nanopub 2012, Micropub 2013). There are also other forms of data publications including, for example, ISA Tools and the *Scientific Data* journal (ISA Tools 2013, Scientific Data 2013).

It is worth noting that multi-omics data from a longitudinal study of a single individual (e.g., the Smarr and Snyder datasets) in their entirety constitute essentially a whole new data type. Supplied with detailed metadata, these data could become a part of a greater, well-documented collage of data within a specific domain. Due to the large amount of data and the complexity in the data acquisition it is exceedingly difficult to capture, disseminate, and interpret the metadata. Generally, minimal reporting requirements are aimed at enabling replication of an experiment, a concept that is not easily applied to the longitudinal personal omics studies. Reuse of data can be enabled with more succinct and concise reporting.

The checklist we propose therefore has a simple structure covering four concise sections: experiment information, experimental design, experimental methods, and data processing (Table 1). The experiment information section includes details of the lab, funding sources, data identification, and a brief abstract to address why the experiment was done. The experimental design section is meant to capture the high level data about the experiment and its statistical design, including sample selection, replication, and randomization. The experimental methods

section contains details about instrumentation and sample preparation. The data processing section captures information regarding methods and tools used in experimental data processing and data analysis.

Table 1.

The metadata captured by this checklist will serve as interlocking bridges for data harmonization and therefore they focus on details of the experimental design and subsequent data analyses. In multi-omics studies, the researcher would fill a checklist for each omics data-type measured. As test cases, two datasets of the integrative personal multi-omics study were used (Snyder, Submitted). The proposed checklist integrates existing ontologies and standards in order to standardize terminology and simplify data input. In its short structured form, the checklist captures important experimental parameters and strikes a balance between comprehensiveness and ease of use. As such, the checklist can serve as a guide to the design of omics studies.

Implementation of this checklist will enable efficient portability and meta-analysis of the data, as well as transparent communication and greater reproducibility of omics studies. Yet the checklist is just the first step towards full utilization of the data. Traditional publication avenues and new data publications, for example, *OMICS Journal of Integrative Biology, Journal of Proteome Research*, *Big Data*, *eLife*, and *Scientific Data* could test and adopt the format to ensure that the crucial information needed to allow data to be harmonized for broader usage is published (OMICS 2013, Journal of Proteome Research 2013, eLife 2013). The assessment of

the metadata quality and the data they accompany could be done through community resources like PubMed Commons (PubMed Commons 2013, Swartz 2013).

Data submissions to single omics databases such as, for example, ArrayExpress and Geo for transcriptomics, or PRIDE and Proteome Exchange for proteomics, would benefit from both additional omics metadata within the given database and robust harmonization with other data-types in other databases (ArrayExpress 2013, Geo 2013, PRIDE 2013, Proteome Exchange 2013). The checklist could also aid submissions to multi-omics databases, data repositories or data clouds. Examples include: data clouds, such as The Open Science Data Cloud, and data repositories, such as Dryad for raw data, and MOPED for processed data (Open Science Data Cloud 2013, Dryad 2013, MOPED 2013). When compatibility and sharing of data and metadata cease to be an issue, a deeper understanding of cells, organisms, and their communities will ensue.

CONCLUSION

The proposed metadata checklist offers a much-needed and balanced approach to bring about data harmonization across omics studies. This is accomplished while also maintaining the flexibility needed to adapt to complex and ever evolving study designs and omics application contexts in the post-genomics era of the life sciences.

ACKNOWLEDGEMENTS

All authors of this publication are members of the Data-Enabled Life Sciences Alliance (DELSA Global). This manuscript has been reviewed and endorsed by the Alliance. Research reported in

this publication was supported by Seattle Children's Research Institute (SCRI), EMC^2, and Intel. This support is very much appreciated. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Moore Foundation, SCRI, EMC^2, or Intel.

AUTHOR DISCLOSURE STATEMENT

No competing financial interests exist.

REFERENCES

ArrayExpress. (2013). www.ebi.ac.uk/arrayexpress

Barrett T, Clark K, Gevorgyan R, Gorelenkov V, Gribov E, Karsch-Mizrachi I, et al. (2012). BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata. Nucleic Acids Res 40, D57–D63.

Bowden M. (2012). The Measured Man. The Atlantic June 13. www.theatlantic.com/magazine/print/2012/0.7/the-measured-man/309018/

Chen R, Mias G, Li-Pook-Than J, Jiang L, Lam H, Chen R, et al. (2012). Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 148, 1293–1307.

Data-Enabled Life Sciences Alliance. (2013a). www.delsaglobal.org.

Data-Enabled Life Sciences Alliance. (2013b). Multi-Omics Metadata Checklist. http://www.delsaglobal.org/news/publications/item/84-checklist

Editorial. (2103). Announcement: Reducing our irreproducibility. Nature 496, 398.

Editorial. (2011). On the table. Nature Genetics 43, 1.

Dryad. (2013). datadryad.org.

eLife. (2013). elife.elifesciences.org.

Field D, Sansone SA, Collis A, Booth T, Dukes P, Gregurick SK, et al. (2009). 'Omics Data Sharing. Science 326, 234-236.

Geo. (2013). www.ncbi.nlm.nih.gov/geo/

ISA tools. (2013). isa-tools.org.

Journal of Proteome Research. (2013). pubs.acs.org/journal/jprobs.

Kettner C, Field D, Sansone SA, Taylor C, Aerts J, Binns N , et al. (2010). Meeting Report from the Second "Minimum Information for Biological and Biomedical Investigations" (MIBBI) workshop. Standards in Genomic Sciences 3, 259–266.

Kolker E, Altinas I, Bourne P, Faris J, Fox G, Frishman D, et al. (2013). In praise of open research measures. Nature 498, 170.

Kolker E and Stewart E. (2012a). Opinion: Data to Knowledge to Action: Introducing DELSA Global, a community initiative to connect experts, share data, and democratize science. The Scientist April 18.

Kolker E, Stewart E, and Ozdemir V. (2012b). DELSA Global for "Big Data" and the Bioeconomy: Catalyzing Collective Innovation. Industrial Biotechnology 8, 176-178.

Kolker E, Stewart E, and Ozdemir V. (2012c). Opportunities and challenges for the life sciences community. OMICS 16, 138–147.

Mias G, Snyder M. (2013). Personal genomes, quantitative dynamic omics and personalized medicine. Quantitative Biology 1, 71–90.

Micropub. (2013). micropub.org

MOPED. (2013). moped.proteinspire.org.

Nanopub. (2013). nanopub.org.

OMICS Journal of Integrative Biology. (2013). www.liebertpub.com/OMI.

Open Science Data Cloud. (2013). www.opensciencedatacloud.org

PRIDE. (2013). www.ebi.ac.uk/pride

Proteome Exchange. (2013). proteomexchange.org.

PubMed commons. (2013). www.ncbi.nlm.nih.gov/pubmedcommons

Reporting Checklist For Life Sciences Articles. (2013). Nature Publishing Group, May-2013. [Accessed: 01-Aug-2013] http://www.nature.com/authors/policies/checklist.pdf.

Scientific Data. (2013). www.nature.com/scientificdata.

Smarr L. (2012). Quantifying your body: A how-to guide from a systems biology perspective. Biotechnology Journal 7, 980-991.

Snyder M, Mias G, Stanberry L, and Kolker E. Metadata checklist for the integrated personal omics study: proteomics and metabolomics experiments. OMICS, submitted.

Stewart E, Smith T, De Souza A, Faris J, Martens L, Mohin S, et al (2013). DELSA Workshop IV: Launching the Quantified Human Initiative. Big Data 1, 187-190.

Swartz, A. (2013). Post-Publication Peer Review Mainstreamed. The Scientist [accessed 2013-10-23]. www.the-scientist.com/?articles.view/articleNo/37969/title/Post-Publication-Peer-Review-Mainstreamed/

Taylor C, Field D, Sansone S, Aerts J, Apweiler R, Ashburner M, et al. (2008). Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. Nature Biotechnology 8, 889-96.

Tennenbaum JD, Sansone S-A, Haendel M. (2013). A sea of standards for omics data: sink or swim? Journal of the American Medical Informatics Association 2013, in press. [accessed 2013-10-17]. http://jamia.bmj.com/content/early/2013/09/27/amiajnl-2013-002066.full.html.

Yilmaz P, Kottmann R, Field D, Knight R, Cole J, Amaral-Zettler L, et al. (2011). Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nature Biotechnology 29, 415–420.