

Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project

Chris F Taylor^{*1,2}, Dawn Field^{2,3}, Susanna-Assunta Sansone^{1,2}, Jan Aerts⁴, Rolf Apweiler¹, Michael Ashburner⁵, Catherine A Ball⁶, Pierre-Alain Binz^{7,8}, Molly Bogue⁹, Tim Booth², Alvis Brazma¹, Ryan R Brinkman¹⁰, Adam Michael Clark¹¹, Eric W Deutsch¹², Oliver Fiehn¹³, Jennifer Fostel¹⁴, Peter Ghazal¹⁵, Frank Gibson¹⁶, Tanya Gray^{2,3}, Graeme Grimes¹⁵, John M Hancock¹⁷, Nigel W Hardy¹⁸, Henning Hermjakob¹, Randall K Julian Jr¹⁹, Matthew Kane²⁰, Carsten Kettner²¹, Christopher Kinsinger²², Eugene Kolker^{23,24}, Martin Kuiper²⁵, Nicolas Le Novère¹, Jim Leebens-Mack²⁶, Suzanna E Lewis²⁷, Phillip Lord¹⁶, Ann-Marie Mallon¹⁷, Nishanth Marthandan²⁸, Hiroshi Masuya²⁹, Ruth McNally³⁰, Alexander Mehrle³¹, Norman Morrison^{2,32}, Sandra Orchard¹, John Quackenbush³³, James M Reecy³⁴, Donald G Robertson³⁵, Philippe Rocca-Serra^{1,36}, Henry Rodriguez²², Heiko Rosenfelder³¹, Javier Santoyo-Lopez¹⁵, Richard H Scheuermann²⁸, Daniel Schober¹, Barry Smith³⁷, Jason Snape³⁸, Christian J Stoeckert Jr³⁹, Keith Tipton⁴⁰, Peter Sterk¹, Andreas Untergasser⁴¹, Jo Vandesompele⁴² & Stefan Wiemann³¹

The Minimum Information for Biological and Biomedical Investigations (MIBBI) project provides a resource for those exploring the range of extant minimum information checklists and fosters coordinated development of such checklists.

To fully understand the context, methods, data and conclusions that pertain to an experiment, one must have access to a range of background information. However, the current diversity of experimental designs and analytical techniques complicates the discovery and evaluation of experimental data; furthermore, the increasing rate of production of those data compounds the problem. Community opinion increasingly favors that a regularized set of the available metadata ('data about the data') pertaining to an experiment^{1,2} be associated with the results, making explicit both the biological and methodological contexts. Many journals and funding agencies now require that authors reporting microarray-based transcriptomics experiments comply with the Minimum Information about a Microarray Experiment (MIAME) checklist³ as a prerequisite for publication⁴⁻⁷. Similarly, minimum information

guidelines for reporting proteomics experiments and describing systems biology models are gaining broader support in their respective database communities^{8,9}; and progress is being made toward the standardization of the reporting of clinical trials in the medical literature¹⁰. Such minimum information checklists promote transparency in experimental reporting, enhance accessibility to data and support effective quality assessment, increasing the general value of a body of work (and the competitiveness of the originators).

Collaborative minimum information checklist development projects for diverse biologically and technologically delineated subject areas are ongoing. A special issue of the journal *OMICS*¹¹ included invited pieces from eight communities supporting minimum information checklist development projects. However, until recently there were no mechanisms for such projects to coordinate their development. Consequently, the full range of checklists can be difficult to establish without intensive searching, and tracking their evolution is nontrivial. Furthermore,

overlaps in scope and arbitrary decisions on wording and substructuring inhibit their use in combination. These issues present difficulties for checklist users, especially those who routinely combine information from several disciplines. Here we explore some of the issues arising from the development of checklists in relative isolation, discuss the potential benefits of greater coordination and describe the mechanisms we have put in place to facilitate such coordination. In summary, we present the MIBBI project (<http://www.mibbi.org/>), which maintains a web-based, freely accessible resource for checklist projects, providing straightforward access to extant checklists (and to complementary data formats, controlled vocabularies, tools and databases), thereby enhancing both transparency and accessibility, as discussed above. MIBBI is managed by representatives of its various participant communities and is fully open to comment from any interested party. Our goal is to facilitate the development of an integrated checklist resource site for the wider bioscience community.

*A list of affiliations appears at the end of the paper.
e-mail: chris.taylor@ebi.ac.uk, dfield@ceh.ac.uk or sansone@ebi.ac.uk

On the need to harmonize minimum information checklists

The current proliferation of documents specifying the minimum information to provide when reporting particular kinds of experimental data has in large part been driven by the advent of a range of so-called 'omics' (and allied) technologies, many of which operate in a high-throughput mode, thereby generating large volumes of data. These documents have been developed independently for the most part, and as a result feature many arbitrary differences in both wording and structure. This greatly complicates the integration of data sets that comply with different minimum information checklists. Increasing appreciation of the potential value accruing to 'secondary use' of data is also a significant factor⁸, reflecting the

general increase in frequency of data-driven (as opposed to hypothesis-driven) investigations in recent years. These trends have together made the need for coordination and harmonization between groups developing data format and reporting standards a critical issue¹⁰. Throughout this document, the words 'standard' and 'standardization' are used to refer only to the regularization of data capture, representation, annotation or reporting, as opposed to best practices for experimental procedures. Specifically, we refer to three kinds of reporting standards: (i) minimum information checklists or guidelines; (ii) formats (syntax); and (iii) controlled vocabularies and ontologies (semantics).

It is clear that checklists should be developed through close consultation with their spon-

soring practitioner communities, but such checklists should also, we believe, be designed to anticipate 'cross-domain' integrative activities. It is unhelpful to confine checklists for the use of particular technologies to a limited set of biologically delineated communities, or to conceive of any such community as being restricted to a particular set of technologies. Consider mass spectrometry, which is used in the study of proteins, metabolites and even to sequence genes; or consider toxicology, which may use any or all of the available 'omics' technologies in pursuit of the greater understanding of the mode of action of a particular compound. Clearly the vistas from any two locations can overlap substantially, so who can claim sole ownership of any part of the scientific landscape? Initiatives such as that to harmonize the

Box 1 Checklist development projects registered with MIBBI

CIMR (<http://msi-workgroups.sourceforge.net/>)

The Metabolomics Standards Initiative's Core Information for Metabolomics Reporting (CIMR) comprises modules for particular aspects of metabolomics workflows; various biological disciplines (for example, microbiology, mammalian biology, plant biology); analytical techniques such as chromatography and NMR; and the use of various statistical techniques.

MIACA (<http://miaca.sourceforge.net/>)

The Minimum Information About a Cellular Assay (MIACA) checklist relates to the perturbation of cells with various classes of molecule, such as small interfering RNA (siRNA) or small chemical compounds. It also provides guidance on the representation of environmental stressors such as temperature shift or starvation, and combinations thereof.

MIAME (<http://www.mged.org/Workgroups/MIAME/miame.html>)

The Microarray and Gene Expression Data Society's well-established Minimum Information About a Microarray Experiment (MIAME) checklist relates to the use of (micro)arrays (most commonly to assay messenger RNA abundance) and analysis of the data generated.

MIAME/Nutr (<http://www.mged.org/Workgroups/rsbi/rsbi.html>)

MIAME/Tox (<http://www.mged.org/Workgroups/rsbi/rsbi.html>)

MIAME/Env (http://nebc.nox.ac.uk/miame/miame_env.html)

MIAME/Plant (<http://miame-plant.sourceforge.net/>)

The MIAME checklist has recently been extended to capture parameters appropriate to nutrigenomics (/Nutr), toxicogenomics (/Tox), environmental biology (/Env) and phytology (/Plant), in each case adding relevant information about the background to the experiment.

MIAPA (<http://www.mibbi.org/index.php/projects/MIAPA>)

The Minimum Information About a Phylogenetic Analysis (MIAPA) checklist relates to the use of software to align biological sequences, and the subsequent use of algorithms to construct phylogenies or cladograms and to draw inferences from them.

MIAPE (<http://www.psidev.info/miape/>)

The Minimum Information About a Proteomics Experiment (MIAPE) checklist comprises modules for reporting the use of and the interpretation of data from various analytical techniques, such as mass spectrometry, gel electrophoresis or liquid chromatography. Modules addressing the description of the biological material under study are planned.

MIARE (<http://www.miare.org/>)

The Minimum Information About an RNA interference Experiment (MIARE) checklist identifies minimal reporting parameters for aspects of high-throughput RNA interference (for example, siRNA and small hairpin RNA) screens, usually in conjunction with cellular assays (compare MIACA checklist, above) and flow cytometry (compare MIFlowCyt checklist, next).

MIFlowCyt (<http://flowcyt.sourceforge.net/>)

The Minimum Information for a Flow Cytometry Experiment (MIFlowCyt) checklist addresses the use of flow cytometry, especially to measure the phenotype and function of cells; information is required about the sample analyzed; the probe, fluorochrome and instrument used; and the analysis of the data collected.

MIGen (<http://www.mibbi.org/index.php/projects/MIGen>)

The Minimum Information about a Genotyping experiment (MIGen) checklist addresses genotyping based on single-nucleotide and microsatellite-repeat polymorphisms, and genetic association and linkage analysis in humans, with special reference to immunology.

MIGS/MIMS (<http://gensc.org/>)

The Minimum Information about a Genome Sequence (MIGS) checklist is an extension of the metadata traditionally captured by the International Nucleotide Sequence Databases (DDBJ, EMBL and GenBank). It captures information relating to nucleic acid sequence, location and sequencing method. The description of habitat is also being extended by means of the tightly integrated Minimum Information About a Metagenomic Sequence/Sample (MIMS) checklist.

(continued)

Box 1 Checklist development projects registered with MIBBI (continued)

MIMIx (<http://www.psdev.info/>)

The Minimum Information required for reporting a Molecular Interaction experiment (MIMIx) checklist includes the identity of molecules that participate in an interaction (with accession number), the methods by which both the interaction and the identity of the participants were established and the role of these molecules in the context of the experiment (as distinct from their biological role).

MIMPP (<http://www.interphenome.org/>)

The Minimum Information for Mouse Phenotyping Procedures relates to the diverse protocols used to characterize the phenotype of a mouse. The checklist addresses both behavioral and physiological traits.

MINI (<http://carmen.org.uk/standards/>)

The Minimum Information about a Neuroscience Investigation (MINI) checklist identifies the minimum information required to report the use of electrophysiology in a neuroscience study.

MIQAS (<http://miqas.sourceforge.net/>)

The Minimum Information for QTLs and Association Studies (MIQAS) checklist relates to the mapping of quantitative trait loci (QTLs) and their association with genetic markers.

MIqPCR (<http://www.rdml.org/>)

The Minimum Information about a Quantitative Polymerase Chain Reaction experiment checklist addresses the minimal reporting parameters for experiments involving quantitative PCR.

MIRIAM (<http://biomodels.net/miriam>)

The Minimum Information Requested In the Annotation of biochemical Models (MIRIAM) checklist offers formal requirements for describing theoretical models of biochemical systems.

MISFISHIE (<http://mged.sourceforge.net/misfishie/>)

The Minimum Information Specification For *In Situ* Hybridization and Immunohistochemistry Experiments (MISFISHIE) checklist²¹ addresses visual interpretation-based tissue gene expression localization experiments, such as those using *in situ* hybridization or immunohistochemistry.

STRENDa (<http://www.strenda.org/>)

The Standards for Reporting Enzymology Data (STRENDa) initiative, along with participants in the biannual ESCEC (Experimental Standard Conditions of Enzyme Characterizations) symposia, maintain a series of checklists addressing the description of enzymatic activity data and the experiments in which these data were collected. These checklists are subject to permanent review by the community involved.

description of 'sample' (the biological source material for a study)¹² or to develop (separable) community-level extensions to shared core standards such as MIAME to better describe domain-specific studies (for example, in environmental biology¹³) are clearly the order of the day. This throws into relief an important division between analytical approaches and the various subdivisions of the biosciences. Checklists that do not span that division will always achieve greater utility because they can be reused more straightforwardly to construct new, made-to-order checklists for a wider range of workflows.

The management of information from experiments (both data and metadata) requires the adoption of reporting standards that ensure transparency and interoperability and that facilitate the integration and exchange of data from different sources. Reporting standards also facilitate the execution of more powerful queries against repositories of experimental data because core information will be regularized and extended information will be supplied in a well characterized manner. This long-term vision will require significant effort and buy-in from a range of scientific communities spread across many nations, but development of some of the kinds of component required to establish such infrastructure is well underway: Functional Genomics Experiment¹⁴ (FuGE) is an object-oriented data model (with an associated XML-based syntactic format) capable

of capturing a wide range of (meta)data in a consistent manner; Reporting Structure for Biological Investigations (RSBI)¹⁵ provides a foundational *lingua franca* for standards projects (described further below) and builds on this to define a simple, but general, tabular format (ISA-TAB)¹⁶ aligned with FuGE; Ontology for Biomedical Investigations (OBI; <http://obi.sourceforge.net/>) is a broad-scope ontology providing a self-compatible set of terms with which to describe a wide range of biological and medical studies; and the Open Biomedical Ontologies (OBO) Foundry¹⁷ (<http://obofoundry.org/>) coordinates the development of a set of 'gold-standard reference ontologies' (including OBI) that can be used in combination because they are based on common principles and, importantly, because procedures have been established to ensure resolution of the conflicts that might arise where ontologies overlap.

Although the primary purpose of minimum information checklists is to guide researchers in reporting their experiments, they can, for the kinds of projects mentioned above, serve a valuable role as key 'use cases', in that they represent the distilled opinion of a particular community on the information that should normally be captured to effectively describe a particular kind of experiment. They therefore provide a realistic scenario with which to test any resource's suitability for use by a community; for example, for software and database

developers to ensure that their products can handle the specified data appropriately; or for instrument vendors to offer checklist-compliant data set export from their instrument management software. It is also likely that journals and funders will adopt some checklists wholesale, incorporating them into their guidance for authors and applicants.

A resource for minimum information checklists: MIBBI

The activities of standardization groups often go unpublished and may not be accessible at all, practically speaking. A common resource for minimum information checklists, coordinated by a group of community representatives from ongoing standardization activities, will help unify the standardization community. It will assist in recruiting participants to ongoing activities and it will help to maintain transparency of process by providing access to project-related information (for example, status, key players and plans). It will also ease the establishment of new initiatives by providing answers to questions such as, "How do we get started?" and, importantly, "How do we make sure we don't reinvent the wheel?". Such an effort will improve communication, knowledge transfer and integration between checklist development projects hailing from different scientific communities and, further, between different kinds of reporting standards projects, ultimately resulting in simplified access to a broad range

Table 1 The composition of each of the 21 checklists visible through the MIBBI Portal

Concept	CIMR ^a	MIACA	MIAME	MIAME/Env	MIAME/Nutr	MIAME/Plant	MIAME/Tox	MIAPA	MIAPE ^a	MIARE	MIFlowCyt	MIGen	MIGS/MIMS	MIMix	MIMPP	MINI	MIQAS	MiqPCR	MIRIAM	MISFISHIE	STREDA	Row totals
Study inputs																						
Study design	■	■	■	■	■	■	■	■	●	■	■	■	■	■	■	■	■	■	■	■	■	17
Generic organism	■	■	■	■	■	■	■	■	●	■	■	■	■	■	■	■	■	■	■	■	■	18
Cells/microbes	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	7
Plant	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	2
Animal	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	3
Mouse	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	1
Human	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	3
Population	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	3
Environmental sample	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	2
Environment/habitat	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	6
<i>In silico</i> model	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	2
Study procedures																						
Organism maintenance	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	1
Animal husbandry	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	4
Cell/microbe culture	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	4
Plant cultivation	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	2
Acclimation	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	1
Preconditioning/pretreatment	■	■	■	■	■	■	■	■	●	■	■	■	■	■	■	■	■	■	■	■	■	7
Organism manipulation	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	4
Assay inputs																						
Generic study input	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	5
Organism part	■	■	■	■	■	■	■	■	●	■	■	■	■	■	■	■	■	■	■	■	■	14
Organism state	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	4
Organism trait	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	6
Biomolecule	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	3
Synthetic analyte	■	■	■	■	■	■	■	■	●	■	■	■	■	■	■	■	■	■	■	■	■	4
Silencing RNA reagent	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	2
Sample collection	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	3
Sample processing	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	12
Sample storage	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	2
Sample transport	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	1
Assay procedures																						
Detection/tagging/staining	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	13
Generic analysis	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	3
Array-based assay	■	■	■	■	■	■	■	■	●	■	■	■	■	■	■	■	■	■	■	■	■	10
Capillary electrophoresis	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	3
Cell phenotyping	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	2
Clinical test/examination	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	3
Column chromatography	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	3
Electrochemical detection	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	1
Electrophysiology mensuration	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	1
Enzyme activity assay	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	1
Flow cytometry	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	2
FTIR spectroscopy	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	1
Gel electrophoresis	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	2
Image acquisition	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	8
Mass spectrometry	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	3
Molecular interaction detection	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	1

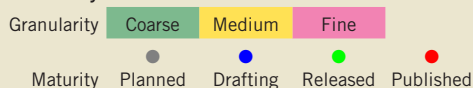
(continued)



Table 1 The composition of each of the 21 checklists visible through the MIBBI Portal (continued)

Concept	CIMR ^a	MIACA	MIAME	MIAME/Env	MIAME/Nutr	MIAME/Plant	MIAME/Tox	MIAPA	MIAPE ^a	MIARE	MIFlowCyt	MIGen	MIGS/MIMS	MIMix	MIMPP	MINI	MIQAS	MIqPCR	MIRIAM	MISFISHIE	STREND ^a	Row totals
Mouse phenotyping																						1
NMR spectroscopy																						2
Nucleic acid sequencing																						3
Quantitative PCR amplification																						1
Toxicology assay																						2
Data analysis																						
Generic data analysis																						11
Clustering																						2
Image characterization																						7
Population genetic analysis																						2
Gene function assignment																						11
Genetic association/mapping																						2
Relative or absolute quantitation																						8
Sequence assembly																						1
Data description																						
Generic data description																						7
Confidence indicator																						4
Data availability																						
Raw data																						12
Processed data																						11
Administrative																						
Citations <i>et cetera</i>																						19
Personnel																						18
Supporting data																						7
Column totals	35	22	14	21	27	19	21	9	21	20	13	24	8	7	11	13	11	6	5	12	12	

Table key



Concepts (row headings) were derived as described in the Methods section, constructed exclusively for the purpose of identifying overlaps between checklists. Because the concepts vary widely in breadth of scope, the number of concepts addressed by any one project is not necessarily indicative of the extent of that project's actual guidelines. Color codings of cells and bullets indicate granularity of coverage and developmental status, respectively. Some bullets have been placed within the matrix itself to provide a finer-grained view of developmental status; cells lacking bullets inherit the overall status of their project (that is, the bullet at that column's head). Row and column totals (counting presence or absence only) are provided in the rightmost column and bottom row. Analyses of these data are provided in **Figure 1** and in **Supplementary Figures 1 and 2**.

^aThe specification is provided as a suite of related documents

of richly annotated data for the end user. Thus, we have established the MIBBI project—a web-based, communal resource designed to act as a 'one-stop shop' for those exploring the range of extant checklist projects and to foster collaborative, integrative development of checklists (<http://www.mibbi.org/>).

MIBBI has two key parts. The first is the 'Portal', which exists simply to raise awareness of, and afford more straightforward access to, a wide range of checklists by providing researchers, journal editors, reviewers, funders and the wider community of checklist developers with a quick and simple way to discover (whether there is) a checklist addressing a particular area and to establish the scope and progress

of the underlying project. The Portal provides summary information for each of the MIBBI-affiliated projects; specifically, the primary contact(s) and website (where available), an overview of the project's scope and developmental status and links to publications and other documents (including, where possible, a link to the most recent version of that project's checklist). Information available through the Portal will be updated as circumstances change (for example, if a project is fragmented or amalgamated, or simply becomes dormant). **Box 1** offers brief textual descriptions of the 21 projects currently registered with MIBBI; **Table 1** provides a representation of the concepts that comprise each project's scope, along

with their checklist's developmental status and, where applicable, an indication that a checklist is composed of separate modules.

By signing up with the MIBBI Portal and thereby attracting more intensive peer oversight, communities will come under pressure to maintain their checklists in light of scientific advances, to provide open access to their processes and to respond to comments. We hope that one of the primary benefits of the Portal will be to raise awareness in the biological and medical communities of the importance of standardization, thereby increasing willingness among researchers to become involved in guiding and shaping the evolution of these activities. We hope it will help push the

community to strive for compliance in their own publication and data-dissemination practices by facilitating access to relevant information about these efforts. We also see this as an excellent artifact with which to promote collaboration within and between communities: the principle we endorse is that if a broadly relevant effort already exists (for example, describing the use of a particular technology), individuals with an interest should seek to join that effort rather than compete with it. However, it is crucial that MIBBI never preclude revisions or innovations; the hoped-for kudos and enhanced coordination accruing to membership should not translate to a possible dominion.

The second key part of MIBBI is the 'Foundry'. Communities can, if motivated, sign up with the Foundry to jointly examine ways to refactor the checklists over which they have control and then to develop a suite of self-consistent, clearly bounded, orthogonal, integrable checklist modules. These modules will then be made available to the community through the MICheckout tool, a collaborative development between the European Bioinformatics Institute and the UK Natural Environmental Research Council's Environmental Bioinformatics Centre. MICheckout will assist users in compiling the correct list of modules and downloading them in a form that they can use. Note that registering a project with MIBBI implies no commitment by a project to participate in

the Foundry activity. Furthermore, attempts to integrate checklists through the Foundry should be managed through a community-driven mechanism that relies primarily on openness and transparency to encourage (voluntary) uptake. The MIBBI Foundry is modeled on the OBO Foundry¹⁷, a newly established initiative in the field of ontology development. Communities working together through MIBBI will produce orthogonal (that is, non-overlapping) minimum information modules, just as the communities involved with the OBO Foundry are aiming to produce orthogonal ontologies.

Foundry activities must be driven by the member communities (acting through their representatives). In preparation for the Foundry activity, we have established discussion forums to facilitate communication between communities to encourage discussion of the overlaps between checklists. Exploratory studies are ongoing, based on coarse comparison tables (such as **Table 1**) that highlight areas addressed by one or more projects. The next stage is to use 'groupware' (that is, a wiki or an online document-sharing tool) to jointly develop modules for those shared areas. Throughout this gradually intensifying activity, we will hold regular face-to-face meetings that act as development workshops and promote good working relationships between project representatives. The first such meeting was held in April 2008 at the European

Bioinformatics Institute and was funded by the UK's Biotechnology and Biological Sciences Research Council. This first meeting rapidly reached consensus on a work plan and established working groups to begin to generate MIBBI Foundry modules. A full workshop report is available through the project's website (<http://www.mibbi.org/>).

High-level abstractions of the components of experimental workflows offer a useful framework to support the integration of checklists. An example of a group attempting to produce such abstractions is the RSBI working group¹⁴, which interacts with a number of other initiatives^{18–20} in working toward an integrated view of functional genomics investigations. In their characterization, an 'Investigation' is a self-contained unit of scientific enquiry, with a holistic hypothesis or objective and a design that is defined by the relationships between one or more 'Studies' and 'Assays'. A Study represents the part of an experiment containing information about the biological material, and an Assay is the part using particular technologies that produce data. The RSBI's proposed framework of well defined, high-level abstractions (such as the three just described) was developed because the above concepts are duplicated, but differently named, across different checklists, confounding the uniform description of the diverse events that may occur within a Study (*sensu* RSBI).

Foundational analysis of MIBBI-registered projects

To better understand the scope and depth of the various MIBBI-registered minimum information checklists, we performed a comparative analysis. **Table 1** presents a projection of the various checklists onto a coarse-grained list of *ad hoc* concepts, constructed exclusively for the purpose of identifying overlaps between those existing checklists; note that the concepts vary widely in breadth of scope (see **Box 2**), so the number of concepts addressed by any one project is not necessarily indicative of the size of that project's guidelines, as some concepts cover whole workflows (for example, 'nucleic acid sequencing'). It will be clear to the reader that some of these concepts, such as 'organism', are almost universal, whereas others, such as 'quantitative PCR amplification', may relate to one group alone. It is also clear that the depth of description required in relation to particular concepts varies widely across projects, suggesting a 'tiered' approach; that is, some of the checklist modules generated by the MIBBI Foundry should, in some cases, require a different depth of description contingent on the particular experimental context. Row and column totals (summing presence or absence only) are

Box 2 MIBBI methods

For the foundational analysis, we created a base data set (**Table 1**) by analyzing the content of the registrant projects' checklists and deriving the list of 65 concepts presented. These concepts were created for the purpose of this analysis and are not taken from any other source, although the meanings of 'study' and 'assay', where they appear, are as set by the RSBI. The concepts have been designed to capture the content of a checklist in an intuitive but compact manner, which means that some concepts represent a large body of methods and technologies (for example, 'nucleic acid sequencing'). However, where a component of such a broad concept was found to have an analog in another project's checklist, that component was factored out to form a new standalone concept (for example, 'detection/tagging/staining' is a concept common to workflows involving microarraying, gel electrophoresis and mass spectrometry), the better to highlight the commonalities between projects. Note also that some concepts are just 'naturally' narrow (such as 'citations *et cetera*', which addresses external referencing, for example, to published papers or to data sets in repositories). The 65 *ad hoc* concepts thus derived have been used throughout the analyses presented here. In some cases, concepts in **Table 1** are indented; this is to indicate that they represent a further specialization of the last less-indented concept above (for example, a 'human' is an 'animal', which is a 'generic organism'). However, the specialization of a more general concept does not imply that those concepts' content overlaps as might be the case in an ontology (that is, 'human' cannot be taken to imply 'animal plus additional information'), and having a specific requirement (for example, 'human') does not imply that there is also generic (that is, for any organism) guidance. The concepts have been represented thus simply to guide the eye while demonstrating that a project may address a concept in a generic or a specific manner, or may actually provide both kinds of requirement (six projects do this, to varying degrees).

provided in **Table 1**; the row totals have been used to rank-order concepts by 'popularity'. **Figure 1** lists the eighteen most common *ad hoc* concepts.

To support greater understanding of the relatedness of the different projects and of the various *ad hoc* concepts, we conducted two pairwise comparisons using the data presented in **Table 1**: concepts 'shared' between pairs of projects, and pairs of concepts occurring together within projects (counting presence or absence only). **Supplementary Figure 1** online illustrates the interrelatedness of the 21 MIBBI-registered projects both as a tree and as an interaction graph. These two representations make clear that there is a subset of closely related (that is, heavily overlapping) projects; these are, broadly speaking, the 'technologically delineated' projects, such as MIAME and the Minimum Information About Proteomics Experiment (MIAPE). It is also clear that there are many projects that are 'related' (according to the tree, if considered in isolation) only by their low degree of relatedness to any other project (as the interaction graph makes explicit). **Supplementary Figure 2** online presents an unrooted tree expressing the relatedness of individual concepts. Although this analysis is based on the various projects' scopes, rather than any sense of the similarity of the concepts themselves, it produces some sensible-looking groupings. All the highly ranked ('high-priority') concepts from **Figure 1** cluster together because most of the projects share an interest in many of them, so they are often found to occur together in individual projects' scopes. Such an analysis can help in deciding how the *ad hoc* concept-based survey presented in **Table 1** should be used as we draft the checklist modules that will ultimately be developed by participants in the MIBBI Foundry's activities (that is, whether some concepts can be combined, whether others should be further subdivided, and so on).

These various analyses make two things plain: first, that there are standout priority areas for the MIBBI Foundry (for example, the uniform description of an organism); and second, that there are many niche areas where little or no collaborative activity is required (for example, the process of mouse phenotyping)—a simple endorsement by MIBBI of the products of a particular project being sufficient, as things stand.

Conclusions

By providing easy access to checklist development projects and their products, MIBBI will facilitate the discovery of checklists appropriate to the needs of practitioners from diverse parts of biological and biomedical science (the

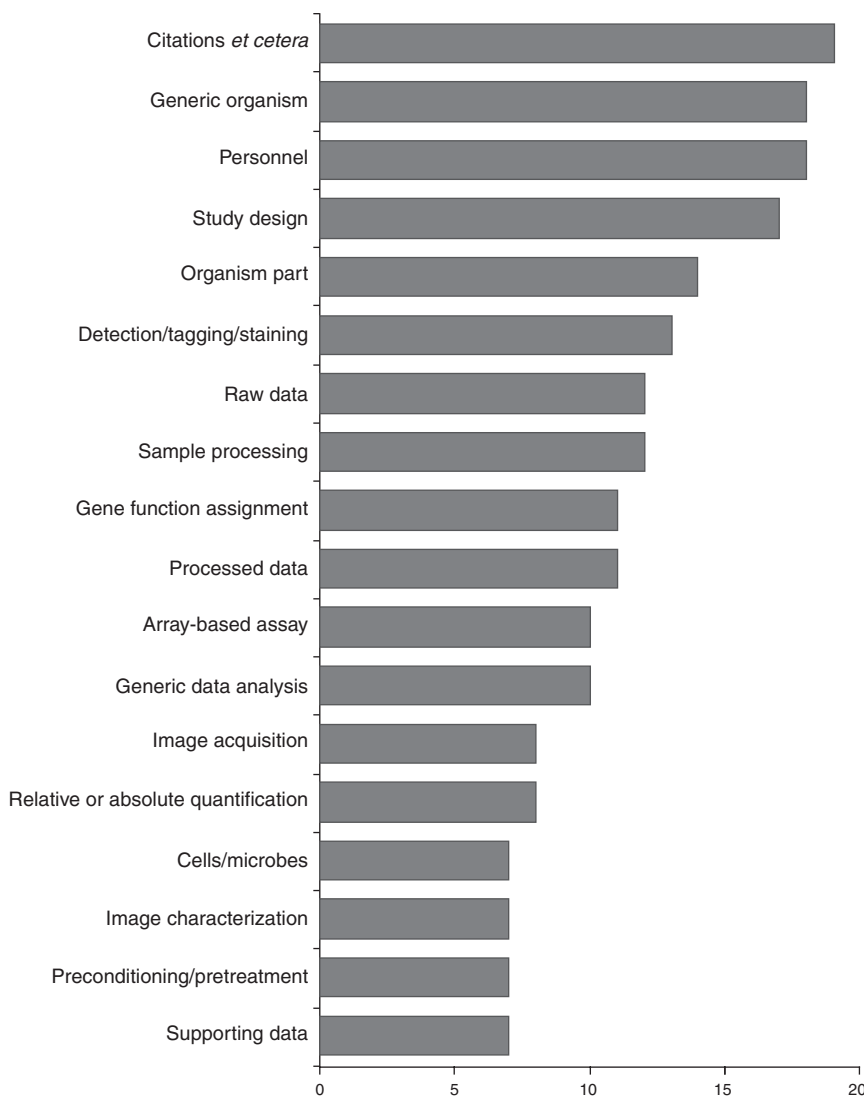


Figure 1 The eighteen highest-ranked *ad hoc* concepts, according to **Table 1**. This highlights priority areas for the MIBBI Foundry (though the concepts used here may not directly translate into guidelines modules).

'one-stop shop' principle). The widespread availability of well-annotated data sets, ensuant to the routine use of minimum information checklists, will increase secondary use of data and allow for a more thorough assessment of the worth of a body of work, making for more efficient and effective science.

MIBBI will increase connectivity between minimum information checklist projects and, more widely, will increase connectivity with projects developing other kinds of informatics resources (formats, vocabularies, tools, databases). The resultant evolution of an interdisciplinary community of checklist developers will bring into focus the collective expertise residing in that group. It will accelerate the establishment of mutually beneficial networks of expertise, and it will advance (through the

MIBBI Foundry, building on the foundational analysis presented here) our long-term vision of a fully integrated, broad-coverage suite of minimum information checklists, in step with the general movement in the biological and medical sciences toward integrated, multifaceted investigations of the puzzles that remain to be addressed in the postgenomic era.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We acknowledge funding from the UK Natural Environment Research Council's Environmental Bioinformatics Centre and the UK Biotechnology and Biological Sciences Research Council (BB/E025080/1) to D.F. and S.-A.S. to support C.F.T. and MIBBI. Work on MIFlowCyt is supported by the US National Institutes of Health's National

Institute of Biomedical Imaging and Bioengineering (EB005034-01) and by Bioinformatics Integration Support Contract A140076 from the US National Institute of Allergy and Infectious Diseases. R.R.B. is supported by the Michael Smith Foundation for Health Research, by the International Society for the Advancement of Cytology and by grant funding from the US National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health (R01EB005034). N.W.H. acknowledges the support of the European Union Framework VI project METAPHOR (Food-ST-2006-03622). F.G., P.L. and work on CARMEN are supported by the UK Engineering and Physical Sciences Research Council (EP/E002331/1). K.T. acknowledges support from Science Foundation Ireland. Work on MIAME/Tox and MIAME/Nutr by P.R.-S. is supported by the NuGO (NoE 503630) and CarcinoGenomics (PL 037712) European Union projects. Work on MIARE is supported by the eDIKT project. Opinions, findings and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the US National Science Foundation or the US National Institutes of Health.

AUTHOR CONTRIBUTIONS

D.F., S.-A.S. and C.F.T. conceived and designed the concept of MIBBI as synergistic project; D.F. and S.-A.S. raised the funds to support MIBBI activities, and

C.F.T. performed the analysis presented in this paper; all authors discussed the results and implications and commented on the manuscript.

1. Quackenbush, J. Standardizing the standards. *Mol. Syst. Biol.* **2**, 2006.0010 (2006).
2. Anonymous. Under the MIAME sun. *Nat. Methods* **3**, 415 (2006).
3. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat. Genet.* **29**, 365–371 (2001).
4. Anonymous. Microarray standards at last. *Nature* **419**, 323 (2002).
5. Ball, C.A. *et al.* A guide to microarray experiments—an open letter to the scientific journals. *Lancet* **360**, 1019 (2002).
6. Ball, C.A. *et al.* Standards for microarray data. *Science* **298**, 539 (2002).
7. Ball, C.A. *et al.* The underlying principles of scientific publication. *Bioinformatics* **18**, 1409 (2002).
8. Taylor, C.F. *et al.* The minimum information about a proteomics experiment (MIAPE). *Nat. Biotechnol.* **25**, 887–893 (2007).
9. Le Novère, N. *et al.* Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.* **23**, 1509–1515 (2005).
10. Altman, D.G., Simera, I., Hoey, J., Moher, D. & Schulz, K. EQUATOR: reporting guidelines for health research. *Lancet* **371**, 1149–1150 (2008).
11. Field, D. & Sansone, S.-A. A special issue on data standards. *OMICS* **10**, 84–93 (2006).
12. Morrison, N. *et al.* Concept of sample in OMICS technology. *OMICS* **10**, 127–137 (2006).
13. Morrison, N. *et al.* Standard annotation of environmental OMICS data: application to the transcriptomics domain. *OMICS* **10**, 172–178 (2006).
14. Jones, A.R. *et al.* The Functional Genomics Experiment model (FuGE): an extensible framework for standards in functional genomics. *Nat. Biotechnol.* **25**, 1127–1133 (2007).
15. Sansone, S.-A. *et al.* A strategy capitalizing on synergies: the Reporting Structure for Biological Investigation (RSBI) working group. *OMICS* **10**, 164–171 (2006).
16. Sansone, S.-A. *et al.* The first RSBI (ISA-TAB) workshop: “Can a simple format work for complex studies?” *OMICS* published online, doi:10.1089/omi.2008.0019 (2008).
17. Smith, B. *et al.* The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* **25**, 1251–1255 (2007).
18. Taylor, C.F. *et al.* The work of the Human Proteome Organisation’s Proteomics Standards Initiative (HUPO PSI). *OMICS* **10**, 145–151 (2006).
19. Fiehn, O. *et al.* Establishing reporting standards for metabolomic and metabonomic studies: a call for participation. *OMICS* **10**, 158–163 (2006).
20. Field, D. *et al.* The minimum information about a genome sequence (MIGS) specification. *Nat. Biotechnol.* **26**, 541–547 (2008).
21. Deutsch, E.W. *et al.* Minimum information specification for *in situ* hybridization and immunohistochemistry experiments (MISFISHIE). *Nat. Biotechnol.* **26**, 305–312 (2008).

¹European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK. ²Natural Environment Research Council Environmental Bioinformatics Centre, Mansfield Road, Oxford OX1 3SR, UK. ³Molecular Evolution and Bioinformatics Group, Oxford Centre for Ecology and Hydrology, Mansfield Road, Oxford OX1 3SR, UK. ⁴Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK. ⁵Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK. ⁶Stanford Microarray Database, Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305-5307, USA. ⁷Swiss Institute of Bioinformatics, Michel-Servet 1, 1211 Geneva, Switzerland. ⁸Geneva Bioinformatics (GeneBio) SA, Avenue de Champel 25, 1206 Geneva, Switzerland. ⁹Jax Mouse Phenome Project, The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609, USA. ¹⁰Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver V5Z 1L3, British Columbia, Canada. ¹¹The Lance Armstrong Foundation, PO Box 161150, Austin, Texas 78716-1150, USA. ¹²Institute for Systems Biology, 1441 N. 34th Street, Seattle, Washington 98103, USA. ¹³University of California Davis, Genome Center, 451 East Health Sciences Drive, Davis, California 95616, USA. ¹⁴National Institute of Environmental Health Sciences and Lockheed Martin Information Technology, Research Triangle Park, North Carolina 27709-2233, USA. ¹⁵Division of Pathway Medicine, University of Edinburgh Medical School, The Chancellor’s Building, Little France Crescent, Edinburgh EH16 4SB, UK. ¹⁶School of Computing Science, Newcastle University, Newcastle upon Tyne NE1 7RU, UK. ¹⁷Bioinformatics Group, Medical Research Council Mammalian Genetics Unit, Harwell, Oxfordshire OX11 ORD, UK. ¹⁸Department of Computer Science, Aberystwyth University, Pengllys, Aberystwyth SY23 3DB, UK. ¹⁹Indigo BioSystems, Inc., 111 Congressional Boulevard, Suite 160, Carmel, Indiana 46032, USA. ²⁰Division of Molecular and Cellular Biosciences, National Science Foundation, 4201 Wilson Boulevard, Arlington, Virginia 22230, USA. ²¹Beilstein-Institut zur Förderung der Chemischen Wissenschaften, Trakehner Strasse 7-9, D-60487 Frankfurt am Main, Germany. ²²Office of Technology and Industrial Relations, Office of the Director, National Cancer Institute, Bldg 31A, Rm 10A52, Bethesda, Maryland 20892, USA. ²³Seattle Children’s Hospital Research Institute, 1900 9th Avenue, Seattle, Washington 98101, USA. ²⁴Seattle Children’s Hospital and Regional Medical Center, 4800 Sand Point Way NE, Seattle, Washington 98105, USA. ²⁵Department of Plant Systems Biology, Vlaams Instituut voor Biotechnologie, Technologiepark 927, B-9052 Ghent, Belgium; Molecular Genetics, Ghent University, B-9052 Ghent, Belgium; and Department of Biology, Norwegian University of Science and Technology, N-7491 Trondheim, Norway. ²⁶Department of Plant Biology, University of Georgia, Athens, Georgia 30602-7271, USA. ²⁷Department of Molecular and Cellular Biology, Life Sciences Addition, University of California, Berkeley, California 94729-3200, USA. ²⁸Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Texas 75390 USA. ²⁹RIKEN Genomic Sciences Center, 3-1-1 Koyadai, Tsukuba-shi, Ibaraki 305-0074 Japan. ³⁰Economic and Social Research Council Centre for Economic and Social Aspects of Genomics (Cesagen), Lancaster University, Institute for Advanced Studies, IAS Building, County South, Lancaster LA1 4YD, UK. ³¹Division of Molecular Genome Analysis, German Cancer Research Center, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany. ³²School of Computer Science, University of Manchester, Oxford Road, Manchester M13 9PL, UK. ³³Department of Biostatistics and Computational Biology and Department of Cancer Biology, Dana-Farber Cancer Institute, Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA. ³⁴Department of Animal Science, Center for Integrated Animal Genomics, Iowa State University, 2255 Kildee Hall, Ames, Iowa 50011-3150, USA. ³⁵Bristol-Myers Squibb, Route 206 & Province Line Road, Princeton, New Jersey 08543-4000, USA. ³⁶NuGO, The European Nutrigenomics Organisation (<http://www.nugo.org/>). ³⁷Department of Philosophy and Center of Excellence in Bioinformatics and Life Sciences, University at Buffalo, New York 14260, USA. ³⁸AstraZeneca UK Ltd., Brixham, Devon TQ5 8BA, UK. ³⁹Department of Genetics, Center for Bioinformatics, 423 Guardian Drive, Philadelphia, Pennsylvania 19104-6021, USA. ⁴⁰School of Biochemistry and Immunology, Trinity College Dublin, College Green, Dublin 2, Ireland. ⁴¹Laboratory of Molecular Biology, Department of Plant Science, Wageningen University, The Netherlands. ⁴²Center for Medical Genetics, Ghent University Hospital, 9000 Ghent, Belgium.