PAS 246:2015

Use of standards for digital biological information in the design, construction and description of a synthetic biological system – Guide

Innovate UK
Technology Strategy Board

bsi.

Publishing and copyright information

ISBN 978 0 580 87166 5

ICS 25.020, 35.240.80

Publication history

First published April 2015

Contents

Foreword

This PAS was sponsored by Innovate UK. Its development was facilitated by BSI Standards Limited and it was published under licence from The British Standards Institution. It came into effect on 30 April 2015.

Acknowledgement is given to Christopher Hirst as the technical author, and the following organizations that were involved in the development of this PAS as members of the steering group:

- Brunel University
- CHAIN Biotechnology Ltd
- Dr Reddy's Laboratories
- Genabler
- GlaxoSmithKline plc
- Imperial College London
- Ingenza Ltd
- Innovate UK
- Newcastle University
- Sphere Fluidics Ltd
- SynbiCITE
- Synthace Ltd
- Visbion Ltd

Acknowledgement is also given to the members of a wider review panel who were consulted in the development of this PAS.

This PAS is not to be regarded as a British Standard. It will be withdrawn upon publication of its content in, or as, a British Standard.

The PAS process enables a guide to be rapidly developed in order to fulfil an immediate need in industry. A PAS may be considered for further development as a British Standard, or constitute part of the UK input into the development of a European or International Standard.

Use of this document

As a guide, this PAS takes the form of guidance and recommendations. It should not be quoted as if it were a specification or a code of practice and claims of compliance cannot be made to it.

Presentational conventions

The guidance in this PAS is presented in roman (i.e. upright) type. Any recommendations are expressed in sentences in which the principal auxiliary verb is "should".

Commentary, explanation and general informative material is presented in italic type, and does not constitute a normative element.

Spelling conforms to The Shorter Oxford English Dictionary. If a word has more than one spelling, the first spelling in the dictionary is used.

Contractual and legal considerations

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

Compliance with a PAS cannot confer immunity from legal obligations.

Innovate UK statement

Innovate UK – the new name for the Technology Strategy Board – is the UK's innovation agency. We fund, support and connect innovative businesses to accelerate sustainable economic growth.

Timely, consensus-based use of standards plays a vital role in ensuring that the knowledge created in the UK's research base is commercialised and brought to market and plays an important part in driving innovation.

Innovate UK is working with BSI, the Research Councils and Catapults to establish new standards earlier in the development of technologies. We are collaborating in four areas of innovation to define standards that will accelerate the development of technologies and services to provide UK businesses with a competitive "first mover advantage", including the subject of this document that will enable computer-aided design, manufacture, and verification using digital biological information.

We have also joined with the Engineering and Physical Sciences Research Council and Biotechnology and Biological Sciences Research Council to create SynbiCITE, a pioneering Innovation and Knowledge Centre dedicated to promoting the adoption and use of synthetic biology by industry. The centre is focused at Imperial College, London and will help turn academia and industry-based research into commercial success. For more information see **http://synbicite.com/**

More widely, health and care is a key priority area in our work – with major innovation programmes to stimulate the development of new technologies, products and services, building on the UK's world-class science and technology base and its global presence in the biopharmaceutical and health technology sectors.

Read more about Innovate UK and our plans in health, care and other areas here: **www.innovateuk.gov.uk** or contact **support@innovateuk.gov.uk**

Introduction

Synthetic biology has been identified by the Synthetic Biology Roadmap [1], Eight Great Technologies report [2] and the Davos Global Risks 2015 report [3] as one of the key emerging technologies with the potential to provide solutions to many of the social, technological and environmental problems of the coming decades. Several areas where immediate attention is required include:

- food security;
- energy security;
- economic security; and
- wealth creation.

Synthetic biology offers a new way to approach such issues. This new approach, based on the use of engineered biology, is likely to be more sustainable than current methods and might offer renewable solutions to such problems.

Synthetic biology is the term given to the conceptual framework surrounding the systematic design and engineering of biological systems. At present the major thrust of effort within synthetic biology is focused on the genetics of biological systems and improving the ability to design and build at the genetic level in order to produce products for a range of applications. The objective of this technology is to enable the formation of a biomanufacturing industry where it will be possible to rapidly design and build new biological systems and organisms for a diverse range of applications. In order for this to happen, improvements to the engineering of biology are required to bring it in line with the rational design used in other engineering disciplines. In particular, this requires the development of the capability to apply the design-build-test cycle to synthetic biological systems and subsequent improvement in the predictive power of the designs. To achieve this, a standards framework is required.

Standards enable effective engineering by establishing both good practice and the necessary framework of tools and processes required to ensure compatibility and interoperability for those working with the technology. At present there are no formalized standards within synthetic biology. There are however community-led projects to standardize various processes within synthetic biology, with particular focus on the transfer of digital biological information between institutions and individuals and the methods for the assembly of genetic constructs. These projects are generally focused on developing openly accessible resources for various groups to use with their own tools and projects. This open nature serves the responsible research and innovation (RRI) framework by providing transparency, evidence and data in the development of technical standards. The adoption and further development of the most useful of these standards will be advantageous for synthetic biology as they will:

- enable the transfer of information between tools and designers;
- improve the reliability and reproducibility of designs; and
- allow the sharing of good practice such as the latest developments in prevention of DNA release to the environment [1)].

In the longer term, formal standardization has the potential to foster the generation of a new digital biomanufacturing industry and accelerate research and innovation within academia.

[1)] MANDELL D. J., LAJOIE M. J., MEE M. T., TAKEUCHI R., KUNZETSIV G., NORVILLE J. E., GREEG C. J., STODDARD B. L. and CHURCH G. M. Biocontainment of genetically modified organisms by synthetic protein design. *Nature.* (2015) and ROVNER A. J., HAIMOVICH A. D., KATZ S. R., LI Z., GROME M. W., GASSAWAY B. M., AMIRAM M., PATEL J. R., GALLAGHER R. R., RINEHART J. and ISAACS F. J. *Recoded organisms engineered to depend on synthetic amino acids. Nature.* (2015).

1 Scope

This PAS gives guidance on the systematic approach to the use of standards for digital biological information in the design, construction and description of a synthetic biological system.

This PAS provides guidance on:

- how digital biological information can benefit the process of designing, constructing and describing biological systems (Clause **3**);
- standards in development for the transfer of digital biological information (Clause **4**);
- metrology requirements for digital biological information (Clause **4**); and
- development of standards for constructing biological systems with digital biological information (Clause **5**).

It does not give specific technical requirements about how to construct such processes, but contains reference to what standards are available and in development, and guidance on their use.

It does not cover:

- specific details of the methods for generating digital biological information;
- how to utilize digital biological information within an existing manufacturing process;
- specific details regarding the design of new manufacturing processes around digital biological information; and
- the utilization of digital biological information and its implementation within the responsible research and innovation (RRI) framework.

It is intended for use by companies, academics and institutions looking to innovate using digital biological information or intending to develop commercially acceptable manufacturing processes.

2 Terms, definitions and abbreviations

For the purposes of this PAS, the following terms and definitions apply.

2.1 Terms and definitions

2.1.1 abstract level

conceptual rather than physical level

2.1.2 biological product

product produced by a living organism

NOTE *Biological products can include, for example, enzymes or chemicals.*

2.1.3 component

DNA object within the synthetic biology open language (SBOL) data model

2.1.4 device

combination of parts that carry out a task

NOTE *This term originated with the BioBrick-based abstraction hierarchy.*

2.1.5 DICOM-SB

data system developed from the existing DICOM standard data model, specifically designed for synthetic biology

NOTE *This incorporates the full range of characterization data (including colour images, where appropriate) and metadata, for a given characterization protocol.*

2.1.6 digital biological information

functional design characteristics of a synthetic biology system

NOTE *This also applies to the end use and includes the behavioural properties, the characteristics of the system implementation and the contexts within the system.*

2.1.7 characterization information

information quantitatively describing the function of a biological part or system

2.1.8 chassis

operating environment for a system

NOTE *If the chassis is a living organism, it may also be called a host.*

2.1.9 host

living organism or cell-like environment in which biological systems operate

NOTE *A host is a type of chassis.*

2.1.10 fragment

linear piece of double-stranded DNA often generated as an intermediary molecule in an assembly process

2.1.11 idempotent assembly method

method of DNA assembly where the sequences required to assemble further pieces are regenerated after addition of each new piece

2.1.12 *in silico*

on a computer

2.1.13 *in vitro*

in solution outside of a living host

2.1.14 *in vivo*

in a living host

2.1.15 interoperability

ability of different parts to be combined and work together consistently

2.1.16 oligos

short DNA sequences often used as primers in polymerase chain reaction (PCR) reaction

2.1.17 part

an element of a device which carries out a single simple function

NOTE *This term originated with the BioBrick-based abstraction hierarchy.*

2.1.18 request for comment (RFC)

community-led system for publishing and modifying standards

2.1.19 specified fragment end

specified restriction overhangs and flanking sequences produced at the end of linear DNA molecules that permit their assembly in accordance with a request for comment (RFC) standard

2.1.20 synthetic biological system

group of biological devices which carry out a human-defined task in a chassis

2.1.21 synthetic biology open language (SBOL)

language and data model for describing designs for synthetic biological systems

NOTE *Includes terms for the components and DNA sequences for genetic design, host, context, in silico models and interactions.*

2.1.22 synthetic circuit

synthetic biological device engineered to perform a human-defined biological operation

2.2 Abbreviations

For the purposes of this PAS, the following abbreviations apply.

BBF	BioBricks Foundation
CAD	computer aided design
CPEC	circular polymerase extension cloning
DICOM	digital imaging and communications in medicine
DICOM-SB	digital imaging and communications in medicine – synthetic biology
DNA	deoxyribonucleic acid
dPCR	digital polymerase chain reaction
MODAL	modular overlap-directed assembly based on linkers
mRNA	messenger RNA
PCR	polymerase chain reaction
qPCR	quantitative polymerase chain reaction
RBS	ribosome binding site
RFC	request for comment
RNA	ribonucleic acid
RRI	responsible research and innovation
RT-qPCR	Reverse Transcriptase quantitative polymerase chain reaction
RPU	relative promoter units
SBML	systems biology markup language
SBOL	synthetic biology open language
SLiCE	seamless ligation cloning extract
USER	uracil-specific excision reagent

3 Design, construction and description using digital biological information

Organisms thrive in a variety of complex and difficult environments. The key to this is their ability to evolve and change to take advantage of their surroundings. This evolutionary process has worked to the advantage of humans for thousands of years, with the best examples of crop and animal species bred to produce desirable traits within that species. The biotechnology developments of the last century have demonstrated that it is possible to use organisms for the production of a wide variety of products, with medically useful proteins, such as insulin, now in common use thanks to their production by microorganisms.

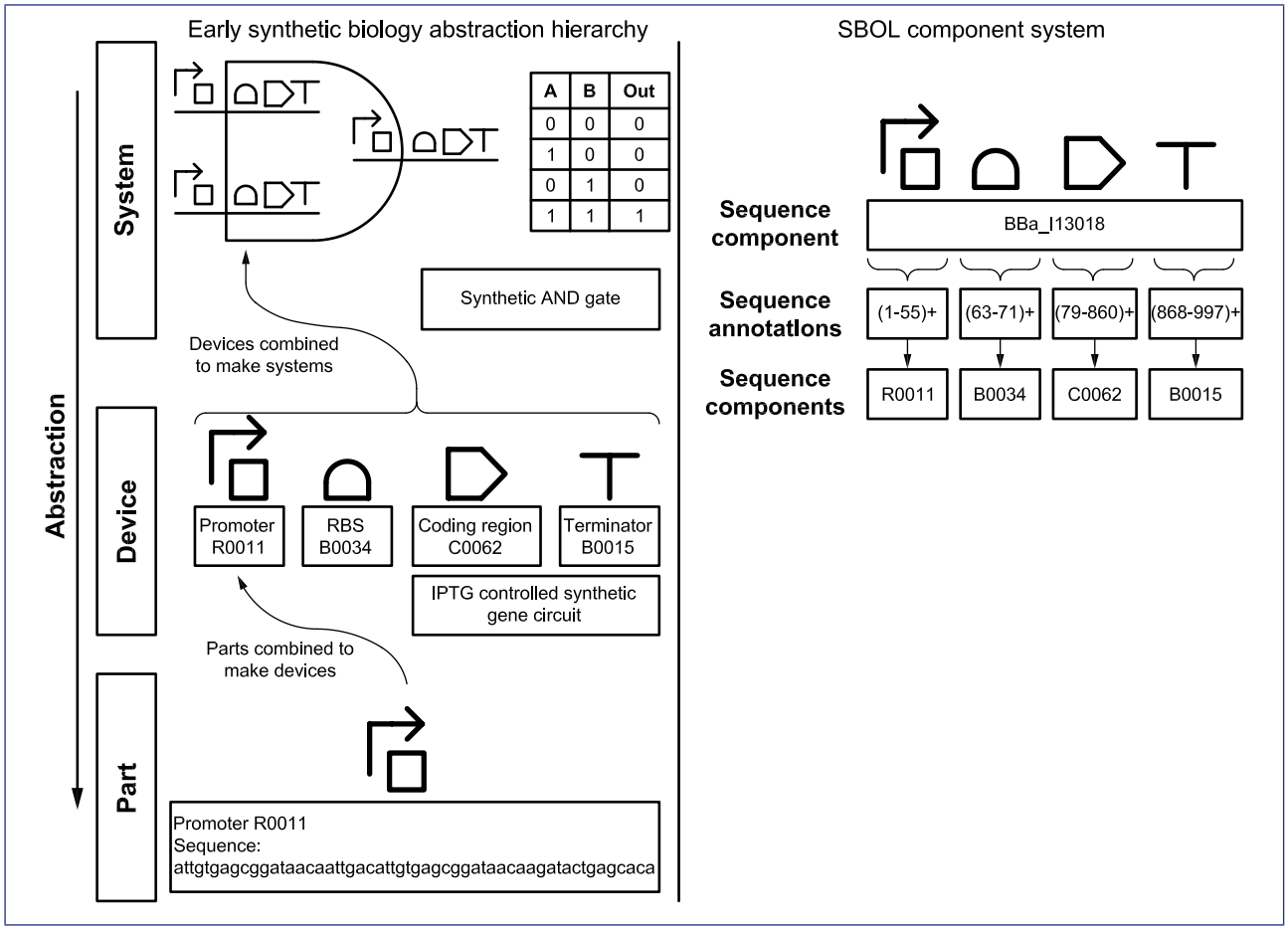
Within the cells of organisms, a large variety of biological molecules interact with each other and any system which operates within such a host. The effect of these interactions between host and system are currently unpredictable. Understanding how to design synthetic systems that operate as expected within the context of these complex organisms will lead to further developments in the synthetic biology field.

Academic and industrial efforts to understand and work with these complex environments have led to increased knowledge around engineering within this context and the development of new technologies to make such engineering easier. The ability to assemble pieces of DNA of increasing size and at increasing levels of scale has enabled synthetic biologists to design increasingly complex and ambitious new devices and systems to operate within a biological chassis. Hand-in-hand with this has been the development of new tools which make it easier to engineer genomes by inserting new sequences or removing unwanted DNA. New strains of organisms have also been developed which are more suitable for engineering, as have methods for reducing some of the dependence of parts of a synthetic biological system or synthetic circuit on the biological context. A range of parts and system design and simulation tools have been developed to take advantage of these developments, and biophysics knowledge has led to the development of calculators, such as the RBS calculator, which predict how certain parts behave based on their sequence. These developments feed into the systematic design framework which underpins the bulk of this currently genetics-based engineering of biology. This framework seeks to incorporate the approach and good practice from other engineering fields with the objective of speeding up developments thorough the use of simulation and accurate characterization.

Just like other engineered systems, synthetic biological systems can be abstracted. An abstraction hierarchy was originally devised for synthetic biology with decreasing complexity, as systems are composed of devices and these devices are themselves composed of parts. From these parts, new complex systems can be engineered from the bottom up to carry out a wide variety of functions. This hierarchy is useful for the field in suggesting what should be designed and characterized in order for increasingly complex devices and systems to be built and has become established in the common nomenclature in the field. In addition to the abstraction hierarchy, a second method for conceptualizing designs has also been developed in the form of the SBOL component model, which was developed primarily for design tools. This format breaks systems' designs down into components which are themselves composed of components and, as a result, is somewhat more flexible than the abstraction hierarchy, while remaining fully compatible (see Figure 1). It has been demonstrated that the design and simulation of some simple systems, such as toggle switches, more accurately predict their behaviour when this framework is adopted.

In practice, the conceptual framework for synthetic biology entails the storage of biological parts as digital biological information which can be efficiently searched and used by various design tools, such as those being developed to use SBOL. This digital biological information contains information on both the physical nature of a part and how it functions. The latter enables such design tools to simulate a design to enable prediction of results in advance of testing. The real benefit from this framework will be observed when the large repositories of pre-tested parts and designs that are common in many industries are developed. With a large number of parts to select from, these tools will be able to simulate a number of designs to choose the set of parts most likely to correctly produce the desired systems. Using digital biological information within this framework will enhance the design-build-test cycle, which will increase the efficiency of the process, and could come together with technological advances such as new biological assembly and measurement techniques to increase the robustness, and reduce the time required to develop new products.

Figure 1 – Description of synthetic biological systems by the original parts, device and systems abstraction hierarchy proposed by Endy and the SBOL component system



4 Design and description standards for digital biological information

4.1 General

While the engineering of synthetic biological systems is becoming increasingly possible, an industrial-scale synthetic biological systems engineering industry is only just beginning to emerge. To support this industry a standards infrastructure is required to aid the development of parts, tools and technologies that will in turn allow the development of manufacturing processes. The first standards required for this infrastructure are for the digital biological information itself, as such standards are vital to the effective sharing of such information which will underpin common tools and manufacturing processes.

The bioinformatics and systems biology fields have already developed a number of community-led standards and ontologies for representing biological data as digital biological information. These range from community-led standards for exchanging genetic level data [Genbank, European Molecular Biology Laboratory format, etc.], through to community-led standards for omics-type data, physiology and biological modelling. Many of these standards can be useful for those wanting to work with digital biological information and a comprehensive list is available online [2)]. In particular, some of these standards have been used to annotate the DNA sequence data generated by various genome projects and the biotechnology industry as a whole. These annotations describe the biological processes associated with a DNA sequence but do not always reflect the abstracted parts useful for design and manufacture. This problem is compounded by some parts possessing multiple functions or activities and parts interacting with a given chassis (although the latter problem is often dealt with by using orthogonal parts).

This digital biological information is stored in a variety of databases depending on the type of information stored. Nucleic Acids Research [3)] keeps an extensive list of the available biological databases and these have the potential to be a useful resource for identifying new part sequences, particularly coding regions. In practice, it is important to treat the information contained within these databases with some care as the quality of these annotations is known to vary and as a community, it would be highly desirable for groups within the bioinformatics, systems biology, synthetic biology and metrology communities to work together to improve the quality of these annotations.

Synthetic biological systems are naturally complex and many internal parts can interact with each other in unknown ways, thereby reducing the predictability of designs. As such, accurate documentation of part features, behaviour and interaction is vital. Any unwanted behaviours or interactions could have significant unintended effects on either another part in the system or the chassis in which the system resides, resulting in undesirable behaviour or properties.

For the effective design and construction of synthetic biological systems, an information standard format suited for the sharing of digital biological information is required. For parts to be used in the construction of synthetic biological systems, standards that tackle the following objectives are needed:

- a common glossary across the field;
- information describing the DNA sequence and components which comprise a design;
- information describing the design of a synthetic biological system at an abstract level;
- information quantitatively describing the function of those components; and
- characterized interactions of various factors, including variation.

[2)] See http://biosharing.org

[3)] See www.oxfordjournals.org/our_journals/nar/database/a/

4.2 Standards for component and DNA sequence digital biological information

As all biological parts are encoded in DNA, a standard information format capable of transmitting the sequence of a part and the necessary annotations is needed. This will ensure that part data can be imported into computer aided design (CAD) tools, correctly incorporated into designs, and the finished designs exported without risk of error or incompatibility.

A community standard called SBOL [4] has been developed by a large working group to allow CAD tools to work together. Rather than annotating a DNA sequence, SBOL has a core data model which specifies all the individual components of a synthetic biological system and then assigns sequences to these where one is provided. This allows devices and systems to be designed even when a required part is currently unavailable. These designs can be transferred via XML file between various tools or individuals without risk of parts or sequences becoming confused. While this is the core of SBOL, various extensions to allow it to document other useful information have been proposed. The most documented of these extensions are designed to cover:

a) the characterization information describing the performance of a design;
b) the description of the chassis the design operates in; and
c) the models which can describe a design or its performance.

In addition, an extension to standardize the representation of parts with diagrams, known as SBOL Visual, is widely used in the synthetic biology community and a package of extensions has recently been released that extends the data model to include many molecular and part interactions which will be useful for CAD tools and in the formation of design simulations.

While DNA sequences from genomic databases can be converted to SBOL format, repositories have been designed for those wanting to design and build with digital biological information and the most prominent of these are given in Annex A. In addition to this, a variety of CAD tools which utilize SBOL have been developed to improve design processes. While many tools have permitted low-level, DNA-sequence-based designs, there is an acknowledgment that high-level, abstract designs are also required as they allow the production of designs without pre-specified parts. High-level designs could be automatically populated with parts later in order to produce a response dictated by a model included with the design. The SBOL 2.0 package, which is currently in development, is designed to enable the production of such high-level designs and the sharing of model-based designs in addition to the lower-level designs currently supported by earlier versions of SBOL.

SBOL should be used as the data standard for the transfer of construct designs and DNA sequence forms of digital biological information. Adopting SBOL as the standard for this transfer has the potential to:

a) facilitate the transfer of high-level designs and model-based designs;
b) ensure accurate transfer of DNA sequences and construct designs between labs, companies and individuals;
c) ensure compatibility between tools designed to work with digital biological information in the form of DNA sequences and constructs, which will in turn:
    1) enable tool developers to more quickly produce new tools and improve existing tools;
    2) allow developers to develop specialized tools rather than end-to-end tools; and
    3) give users more freedom in choosing combinations of tools for design rather than end-to-end tools;
d) be used as a standard descriptor in journals/ publications;
e) be used as a standard for teaching the next generation of professionals to ensure long-lasting impact; and
f) add to a body of knowledge which is open and transparent.

4.3 Standards for functional characterizations

4.3.1 General

SBOL can make the design process much simpler, but to get the greatest benefit from using digital biological information within the synthetic biological system engineering process, information regarding the function of parts of these designs is required in order to test designs *in silico*. Simulation prior to assembly and testing allows pre-screening of designs to focus resources on a smaller number of designs which are more likely to behave as intended.

Unlike component and DNA sequence forms of digital biological information, functional information is not uniform. Different biological parts and systems can be described in different ways, and different approaches can be taken to expressing this information. Some parts are known to exhibit multiple behaviours which require description. In addition to the lack of uniformity, the complexity of synthetic biological systems is a significant challenge which is made more difficult by the interaction of synthetic biological systems with various factors. There are three layers to these interactions:

a) the interaction of parts within the system with each other;

b) the interaction of parts with the chassis; and

c) the interactions between chassis within multicellular systems.

Due to the integrated nature of biology, it is not possible to separate the designed and broader system components within the chassis. This is also true for the parts within the system which may interact in unpredictable ways. While there has not been a specific standardization drive to prevent these interactions, strategies to obtain more predictable function have been developed which achieve this via functional isolation. The ability of a part or design to be context-independent is highly desirable and will be the result of greater understanding of part function and interaction with context. This understanding will further improve the predictability of part function, but this will only be achievable with the development of additional metrology suitable for synthetic biology.

4.3.2 Metrology

Metrology is an important tool to aid the design, understanding and development processes within an engineering discipline. Metrological standards provide a framework for the consistent collection and interpretation of results and compliance with such standards increases the ability to compare data. This is particularly useful where large datasets are produced and where big data techniques are used to identify previously unknown information. For synthetic biology, metrology standards will enable the design aspect of the approach which relies heavily on simulation (which in turn is based on accurate quantitative knowledge of part behaviour). It is currently difficult to obtain quantitative descriptions of part behaviour which match the parameters demanded by the models which underpin the design simulations. Suitable metrologies can bridge this gap and improve both the quality and accuracy of simulations and the representativeness of the data generated by characterizations. This in turn would make designs more reliable and has the potential to help increase understanding of the context within systems.

A variety of metrology standards are required to get the most out of synthetic biology. Synthetic biological systems operate inside a chassis which is often a living organism and most measurements are either performed on populations of cells or with purified components *in vitro*. Methods to measure individual cells and activity *in vivo* would be highly advantageous as they would avoid both variations within the population and the artificial *in vitro* context which might not reflect the context in which the part or system operates. For other measurements, metrology based on the concentration of metabolites and molecules or using comparative measures might be more appropriate. The former will be useful for optimization of large, complex systems where monitoring the behaviour of all the parts within the system is impractical, while the comparative measures are likely to be particularly useful where part behaviour is difficult to measure directly or where the final context is less well understood. The data produced according to these frameworks would ideally be dynamic and multi-layered in nature and, where possible, would enable the identification and quantification of more than a single molecule.

The key to appropriate methodology will be its use in the design process and particularly in the simulation of those designs, as improvements to these processes will benefit synthetic biology the most. As a reference guide, the abstraction hierarchy may prove particularly useful as it can be used to suggest the levels where metrology is required. Metrology is required for parts, the more complex devices, e.g. protein expression and function, and also the operation of the system. Data collection methodologies and metrologies that can capture this information will be powerful tools. Some examples of key targets for metrology include:

a) single cell assays;
b) comparative measures;
c) protein metrology including:
    1) concentration and isoform information;
    2) presence and type or post-translation modification; and
    3) activity *in vivo*;
d) concentration of metabolites;
e) biological interactions, e.g. protein-metabolite, protein-protein, or aptamer-metabolite;
f) batch comparison for *in vitro* reaction mixes;
g) DNA metrology, including:
    1) copy number of plasmids and regions of the genome;
    2) topology and biophysics (including bending and solenoid formation); and
    3) epigenetic modifications;
h) RNA metrology, including:
    1) concentration; and
    2) ribozyme activity *in vivo*.

Metrology of biological objects (for example a metabolite or a protein) and activities is an ongoing area of study for the field and those in the metrology community. Some measurement methods and reference materials suitable for the identification of metabolites and biological molecules and quantification of concentration have been developed. Examples of these include the application of mass spectrometry [5] for metabolite and protein quantification and evaluation and the use of quantitative PCR and digital PCR for the quantification of DNA copy number level [6] and Reverse Transcriptase quantitative PCR (RT-qPCR) for RNA transcript levels [7]. Some metrology for biological processes has also been developed by the synthetic biology community with a focus on transcription and protein expression. This includes absolute level metrology for fluorescent protein expression, the rate of transcription initiation in polymerases per second (PoPS) [8] and a bead-based data calibration method to allow comparison of flow cytometry data [9]. A comparative level metrology has also been introduced in the form of relative promoter units (RPU) [10]. Details of these methods are given in Annex B. Ultimately for the further development of metrology, the ideal scenario will be a community-led effort with the help of metrologists imbedded within synthetic biology groups.

Development of standards for metrology and composition that are useful and informative for design is underway by many communities within the science and engineering disciplines and, when complete, will improve functional characterizations by:

i) suggesting suitable quantitative methodologies for the collection of information regarding biological function;
ii) ensuring the information is appropriate for a given part or component;
iii) increasing the degree to which parts, components and systems are functionally isolated during both measurement and operation;
iv) understanding part and subsystem responses when they are not functionally isolated;
v) improving the accuracy and comparability of such information; and
vi) ensuring the data outputs are generated in a form which are compatible with system modelling, simulation and design tools.

4.3.3 Standards for the exchange of characterization information relating to function

In the absence of metrology standards, attention has focused on the exchange of characterization information. There are some community standards for data exchange from the systems biology and bioinformatics fields which can be useful for documenting biological experiments and exchanging data, particularly for omics experiments [11]. Of particular interest to those wanting to share digital biological information are efforts to define the minimal information required to describe datasets by research groups worldwide. The community-led Minimal Information for Biological and Biomedical Investigations (MIBBI) project [4] has worked to collate and promote integration of these reporting guidelines for many experimental methodologies and these guidelines may provide a suitable starting point for documentation requirements.

Within the synthetic biology community an exchange format is in development by a group working within the Flowers Consortium [12] and ST-Flow [13] projects based on the established DICOM (digital imaging and communications in medicine) standard. DICOM is used globally in medical equipment for the sharing of test results and associated metadata for medical imaging, and an extension of this, DICOM-SB (digital imaging and communications in medicine – synthetic biology), is being prepared for functional characterization. DICOM-SB extends the imaging modality data model within DICOM to enable it to describe many methodologies used by synthetic biologists, such as fluorescent assays and flow cytometry, and is being updated as new methodologies are adopted by the field. A single DICOM-SB document describes a characterization experiment carried out on one of a number of parts built into the data model. Like SBOL, DICOM possesses extensions and one of these (DICOM structured reporting) enables the production of datasheets and similar human readable outputs.

The imaging modality data model is important because this stores metadata relevant to understanding the measurements carried out and, using the modality system, this is tailored to each assay used. Storage of the metadata with the characterization information is important because synthetic biological systems operate in a chassis which itself is affected by its environment. This metadata comprises equipment settings, experimental conditions (such as reaction media) and details about the protocol which may impact on the results. For certain datasets, DICOM-SB also provides an efficient storage format, notably flow cytometry and image data types.

As with SBOL, use of DICOM-SB should enable the transfer and citation of functional characterization information in a reliable manner. It might also help tool designers cope with the variety of characterization information by standardizing the format for this information. While SBOL and DICOM-SB are different tools for transferring digital biological information, they are not necessarily competing, as they document different aspects of digital biological information and are regarded by both development teams as compatible. It would be possible for SBOL extensions to contain links to DICOM-SB documents that describe the characteristics of the parts within an SBOL design and vice versa. The two conceptual frameworks may also be able to work in a complementary manner, with DICOM-SB documents providing parameters to SBOL models and SBOL providing contextual data not covered in the DICOM-SB data model. The key to this relationship will be a standard interface layer that allows linking of both documents and standardizing information relating to the chassis and design characterization. Development of the host context and performance SBOL extensions might permit this.

[4] See http://mibbi.sourceforge.net/portal.shtml

5 Construction of biological systems using digital biological information

5.1 General

Following the design of a biological system, it is important to consider the methods for assembling the synthetic biological system. Building large DNA molecules that produce biological products or systems used to be a skilled and time-consuming process, but recently this step has become much simpler due to a proliferation of assembly methods and a reduction in DNA synthesis costs. The majority of assembly methods fall into one of two types:

a) standardized idempotent assembly methods; or

b) flexible scarless methods.

Descriptions of a number of commonly-used assembly methods are given in Annex C.

Rather than choosing a standard assembly procedure it is advisable to select the most appropriate assembly procedure for the task in hand. It is also worth considering that synthesizing a section or all of a design may be the fastest and even cheapest method after labour and reagent costs are factored in.

While standardizing around a particular method or even type of method would not be advisable at present, it is important that there is consistency of assembly processes with regard to the documentation of the resulting construct and the version history and origins of the components used. Given the complex nature of synthetic biological systems and their interactions, the full sequence of all constructs should be documented in full as they are important to the context of the synthetic biological system. Accurately documenting the full sequence, the validation of the sequence, the quality of the annotation, the parts within the system, and any scars between them will help the synthetic biology community as a whole by:

- improving the reproducibility of experimental data; and
- aiding the comparison of datasets and contexts to allow the identification of interactions inside synthetic biological systems and the context in which they reside.

5.2 Idempotent assembly methods

There have been some attempts to develop standards for idempotent assembly methods, with early attempts such as BioBricks™ [5] (BBF RFC 10(2)) [14] and BglBricks (BBF RFC 21(3)) [15] through the BioBrick Foundation's request for comments (RFC) system (see Annex C). These methods are designed to be idempotent as they produce some form of specified fragment ends and use restriction enzymes that cut inside the prefix and suffix or nearby to produce specified overhangs. These idempotent assembly methods have the advantage that:

a) upon completion of an assembly step the produced fragment can immediately be used in the next assembly stage;

b) the standard ends produced make this method more efficient for the assembly of systems where certain parts are varied; and

c) polymerase chain reaction (PCR) is generally not required, making it amenable to automation.

However, there are drawbacks associated with the use of specified fragment ends:

1) they leave scar sequences inside the assembled DNA sequence which may interfere with the activity of adjacent parts; and

2) as the ends are always the same, a limited number of pieces can be assembled at a time (as few as two) and this increases the assembly time as the size of the required construct increases.

[5] Biobrick is the trade mark of a product supplied by the BioBrick Foundation. This information is given for the convenience of users of this document and does not constitute an endorsement by BSI of the product named. Equivalent products may be used if they can be shown to lead to the same results.

5.3 Scarless assembly methods

The alternative to the use of specified fragment ends is to use more flexible assembly methods such as homology-based methods or non-standardized Golden Gate assembly (see Table C.1). These methods are more flexible as the assembly procedure can be designed in such a way that the sequences responsible for joining fragments of DNA will be inside the section being assembled. A variety of homology-based methods exist which use either recombination or the generation of single-stranded DNA to carry out the assembly process. The flexible methods have the advantage that:

a) they can join pieces of DNA without leaving scar sequences which could alter how the parts within the synthetic biological system behave;

b) many of these methods allow multiple pieces to be combined simultaneously and so require significantly less time for the production of large assemblies; and

c) it might be possible to incorporate some short functional DNA sequences within the homology regions to reduce the number of fragments in an assembly.

However, these methods tend to require the use of PCR to add new ends to each fragment in the assembly which, as a result:

1) makes it more difficult and expensive to carry out combinatorial assembly;

2) increases the risk of mutation, thus requiring more careful verification; and

3) makes it significantly more difficult to automate reliably.

5.4 Alternative assembly methods

In addition to these assembly methods, a number of hybrid methods have also been developed which may offer an alternative assembly solution, depending upon the project. For example, MODAL (modular overlap-directed assembly based on linkers) uses specified prefix and suffix sequences for the binding of the oligos which introduce homology sequences, while another technique, termed the Unique Nucleotide System, uses standardized homology sequences to allow combinatorial and large-scale assembly in a similar way (see Annex C). As a result these methods are useful for combinatorial assembly and can reduce some of the homology-associated costs but require the insertion of scar sequences between parts in the assembly.

Annex A (informative)

Repositories of digital biological information

The most prominent repositories for those wanting to design and build with digital biological information are given in Table A.1.

Table A.1 – Synthetic biology repositories of digital biological information

Repository	Notes	Information type
Joint BioEnergy Institute (JBEI)	DNA sequence repository with public and private collections. Linked to the DeviceEditor CAD tool and j5 assembly design software	SBOL
Parts registry	Part repository for the International Genetically Engineered Machine (iGEM) competition. All parts designed to work with the BioBricks™ (BBF RFC 10) assembly method	SBOL
Virtual Parts Repository	Registry of DNA sequences and models for parts. Contains a large number of *Bacillus subtilis* parts and increasingly being populated with *E. coli* and modular model based parts	SBOL and SBML (systems biology markup language)
SEVA (Standard European Vector Architecture) platform	Repository of plasmids designed to be compatible with the SEVA architecture as part of the ST-Flow project. SEVA plasmids have standardized modules and intervening junctions	GenBank although switching to SBOL
SynBIS	A web-based synthetic biology information system which includes a registry of parts (including characterizations and metadata) that is currently in development	DICOM-SB and SBOL

Annex B (informative)

Measurement methods and community-led metrology standards

Details for methods which may be suitable for synthetic biology and community-led metrology development are given in Table B.1.

Table B.1 – Details of methods and metrology which may be useful for synthetic biologists

Measurement method or unit of metrology	Target	Notes
Mass spectrometry methods	Metabolites, small molecules and proteins	Prepared samples can be analyzed by mass spectrometry alone or in combination with chromatography equipment such as liquid phase chromatography. The mass spectrometer identifies compounds based on the mass to charge ratio of ions generated during the procedure. It is possible to use these ratios to identify the corresponding molecule and then the amount of events observed to quantify the concentration of the compound or molecule if appropriate reference materials and controls are used
Reverse Transcriptase quantitative PCR (RT-qPCR)	mRNA transcript levels	Technique where fluorescent dye is used to measure the amplification of DNA in a PCR reaction. Prior to the qPCR reaction, a DNA template is generated from RNA samples using a Reverse Transcriptase enzyme. This DNA is then amplified using suitable primers and the amplification monitored to enable quantification of the mRNA in the original sample
Quantitative PCR (qPCR) and digital PCR (dPCR)	DNA copy number and number of genome insertions	Quantitative PCR uses fluorescent dye to measure the amplification of DNA in PCR reactions. When a qPCR reaction is carried out with appropriate genomic controls and reference material, this can be used to quantify the copy number of a plasmid or genome insertions. Digital PCR uses the same method, however prior to the PCR reaction the sample is diluted into reaction chambers and the number of chambers in which a product is generated by the amplification is used to calculate the quantity of DNA in the samples. Because of this dPCR is likely to be more accurate for gene copy number than qPCR and standards for using dPCR for gene copy number quantification are in development

Table B.1 – Details of methods and metrology which may be useful for synthetic biologists *(continued)*

Measurement method or unit of metrology	Target	Notes
Relative Promoter Units (RPU) or Relative Expression Units	Transcriptional activity	Generally collected from fluorescent protein expression data. Effectively sets the activity for a reference part to a value of one on a scale and scales the observed results relative to it. This requires the use of a reference construct – usually the same construct carrying the J23101 promoter – which can be used to normalize the output of the other samples tested under the same context
Polymerases per second (PoPs)	Transcriptional activity	Collected from fluorescent protein expression data. Requires multiple layers of calibration but in theory gives an absolute representation of transcriptional output. These calibrations include fluorescence to molecules of fluorophore, absorbance to number of cells and the number of fluorophore molecules generated per mRNA transcript
Equivalent units of fluorescent dye	Protein concentration (fluorescent or fluorophore tagged)	Standardization and calibration method originally demonstrated for flow cytometry. This requires fluorescent calibration beads of known dye concentration to be measured as part of the measurement procedure. The results for the fluorescent beads can be used to generate a calibration curve for fluorescence to dye which can then be applied to the results of the samples in the experiment. The resulting data is primarily useful for lab-to-lab comparison

Annex C (informative)

Assembly methods commonly used for construction of biological systems

The most commonly used methods for assembling synthetic biological systems are given in Table C.1.

Table C.1 – Assembly methods for construction of synthetic biological systems

Type	Notes/Type	Type of scar generated	Idempotent
BioBrick™	Standard prefix and suffix-based restriction enzyme cloning	8bp scar which includes stop codon for in frame protein	Yes
BglBrick	Standard prefix and suffix-based restriction cloning	8bp scar more suitable for protein fusions	Yes
MoClo (Specified end version of Golden Gate)	Restriction-based using Type IIs restriction enzymes to generate specified 4 base pair overhangs to join fragments. The specified overhangs relate to the parts they are connected to or their position relative to over expression units	4bp scars between joined fragments	Yes
Golden Braid (Specified end version of Golden Gate)	Restriction-based using Type IIs restriction enzymes to generate specified 4 base pair overhangs to join fragments. The specified overhangs relate to the assembly "level" in use	4bp scars between joined fragments	Yes
Golden Gate	Restriction-based using Type IIs restriction enzymes to generate specifiable 4 base pair overhangs to join fragments	Potentially scarless	Not without specified fragment end
Isothermal assembly	Homology-based DNA assembly. Uses a 5' exonuclease to generate single-stranded DNA. Complementary single-stranded sections of the DNA fragments are then allowed to anneal before they are repaired by DNA polymerase and ligase enzymes	Potentially scarless	Not without specified fragment ends
CPEC	Homology-based DNA assembly. The full sequence is assembled by PCR where the fragments act as primers for each other	Potentially scarless	No

Table C.1 – Assembly methods for construction of synthetic biological systems *(continued)*

Type	Notes/Type	Type of scar generated	Idempotent
In Fusion™	Proprietary homology-based DNA assembly method from Clontech. Uses an enzyme mix to carry out recombination at homology sites	Potentially scarless	No
SLiCE (seamless ligation cloning extract)	Homology-based DNA assembly method. DNA fragments with homologous ends are incubated in cell extract to assemble DNA	Potentially scarless	No
USER™ (uracil-specific excision reagent)	Homology-based DNA assembly. Uses uracils in the DNA sequence to direct the exonuclease activity of some DNA polymerases, leading to formation of single-stranded DNA at the ends of DNA fragments. These single strands are then allowed to anneal before ligation	Potentially scarless	No
Ligase cycling reaction	Uses a single-stranded DNA oligo homologous to two DNA fragments to bridge the gap between the fragments and allow their ligation together. Multiple cycles with multiple single-stranded DNA oligos enables the assembly of many fragments	Potentially scarless	No
MODAL	Uses prefix and suffix sequences to act as attachment sites for linkers. These linkers can then be used as priming sites for PCR reactions to add homologous sequences to the DNA for assembly via CPEC or Isothermal assembly	75bp scar containing 15bp prefix and suffix sequences	No
Paperclip	Homology-based DNA assembly. This uses two pairs of annealed oligos which are ligated together to form a “clip”. Each pair is homologous to one section of DNA in a reaction and so the clip binds both pieces of DNA, bridging them. Methods such as USER, CPEC or SLiCE can then be used to assemble the DNA into a plasmid	3bp scar between joined fragments	No

Table C.1 – Assembly methods for construction of synthetic biological systems *(continued)*

Type	Notes/Type	Type of scar generated	Idempotent
Unique nucleotide system	Homology-based DNA assembly. Uses isothermal assembly to assemble DNA pieces flanked by unique nucleotide sequence linkers	Scar equal in length to the unique nucleotide linker used	No
Yeast assembly	Homology-based DNA assembly. DNA fragments which have overlapping homology sequences are transformed into a yeast cell which recombines the DNA at the homology sites and assembles it into a plasmid	Potentially scarless	No

Bibliography

Referenced publications

[1] TECHNOLOGY STRATEGY BOARD. *Synthetic Biology Roadmap for the UK.* Swindon: TSB, 2012.

[2] WILLETS, D. *Eight Great Technologies.* London: Policy Exchange, 2013.

[3] WORLD ECONOMIC FORUM. *Global Risks 2015 10th Edition.* Geneva: WEF, 2015.

[4] GALDZICKI, M., CLANCY, K. P., OBERORTNER, E., POCOCK, M., QUINN, J. Y., RODRIGUEZ, C. A., ROEHNER, N., WILSON, M. L., ADAM, L., ANDERSON, J. C., BARTLEY, B. A., BEAL, J., CHANDRAN, D., CHEN, J., DENSMORE, D., ENDY, D., GRUNBERG, R., HALLINAN, J., HILLSON, N. J., JOHNSON, J. D., KUCHINSKY, A., LUX, M., MISIRLI, G., PECCOUD, J., PLAHAR, H. A., SIRIN, E., STAN, GB., VILLALOBOS, A., WIPAT, A., GENNARI, J. H., MYERS, C. J. and SAURO, H. M. The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *In: Nature Biotechology*, 2014, Volume 32, no. 6, p. 545–550.

[5] SARGENT, M. (Ed.). Guide to achieving reliable quantitative LC-MS measurements, RSC Analytical Methods Committee. 2013. ISBN 978-0-948926-27-3

[6] WHALE, A. S., HUGGETT, J. F., COWEN, S., SPEIRS, V., SHAW, J., ELLISON, S., FOY, C. A. and SCOTT, D. J. Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. *In: Nucleic Acids Research*, 2012, Volume 40, no. 11.

[7] DEVONSHIRE, A. S., SANDERS, R., WILKES, T. M., TAYLOR, M. S., FOY, C. A. and HUGGETT, J. F. Application of next generation qPCR and sequencing platforms to mRNA biomarker analysis. *In: Methods*, 2013, Volume59, no. 1, p. 89-100.

[8] CANTON, B., LABNO, A. and ENDY, D. Refinement and standardization of synthetic biological parts and devices. *In: Nature Biotechnology,* 2008, Vol. 26, p. 787-793.

[9] BEAL, J., WEISS, R., YAMAN, F., DAVIDSOHN, N. and ADLER, A. A Method for Fast, High-Precision Characterization of Synthetic Biology Devices. *In: MIT CSAIL Technical Reports,* 2012.

[10] KELLY, J. R., RUBIN, A. J., DAVIS, J. H., AJO-FRANKLIN, C. M., CUMBERS, J., CZAR, M. J., DE MORA, K., GLIEBERMAN, A. L., MONIE, D. D. and ENDY, D. Measuring the activity of BioBrick promoters using an in vivo reference standard. *In: Journal of Biological Engineering,* 2009, Volume 3, no. 4.

[11] CHERVITZ, S. A., DEUTSCH, E. W., FIELD, D., PARKINSON, H., QUACKENBUSH, J., ROCCA-SERRA, P., SANSONE, S-A., STOECKERT, C. J., TAYLOR, C. F., TAYLOR, R. and BALL, C. A. Data Standards for Omics Data: The Basis of Data Sharing and Reuse. *In: Methods in Molecular Biology*, 2011, Volume 719, p. 31-69.

[12] FLOWERS CONSORTIUM *Flowers Consortium Homepage* [online]. 2013. Available from: www.synbiuk.org

[13] ST-Flow Available from: www.cnb.csic.es/~stflow-project/ST-Flow/Welcome.html

[14] KNIGHT, T. *BBF RFC 10: Draft for BioBrick Biological Parts* [online]. 2007. BioBrick Foundation. Available from: http://hdl.handle.net/1721.1/45138

[15] ANDERSON, J. C., DUEBER, J. E., LEGUIA, M., Wu, G. C., GOLER, J. A., ARKIN, A. P. and KEASLING, J. D. *BBF RFC 21: BglBricks Assembly Standard* [online]. 2009. BioBrick Foundation. Available from: http://hdl.handle.net/1721.1/46747

Further reading

KITNEY, R. and FREEMONT, P. Synthetic biology – the state of play, *In: FEBS Letters*, 2012, volume 586, no.15, p. 2029-2036.

HUGGETT, J. F., FOY, A. C., BENES, V., EMSLIE, K., GARSON, J. A., HAYNES, R., HELLEMANS, J., KUBISTA, M., MUELLER, R. D., NOLAN, T., PFAFFL, M. W., SHIPLEY, G. L., VANDESOMPELE, J., WITTWER, C. T. and BUSTIN, S. A. The digital MIQE guidelines: Minimum Information for Publication of Quantitative Digital PCR Experiments. *In: Clinical Chemistry*, 2013, Volume 59, no. 6, p. 892-902.

SANSONE, A-S., ROCCA-SERRA, P., FIELD, D., MAGUIRE, E., TAYLOR, C., HOFMANN, O., FANG, H., NEUMANN, S., TONG, W., AMARAL-ZETTLER, L., BEGLEY, K., BOOTH, T., BOUGUELERET, L., BURNS, G., CHAPMAN, B., CLARK, T., COLEMAN, L-A., COPELAND, J., DAS, S., DE DARUVAR, A., DE MATOS, P., DIX, I., EDMUNDS, S., EVELO, C. T., FORSTER, M. J., GAUDET, P., GILBERT, J., GOBLE, C., GRIFFIN, J. L., JACOB, D., KLEINJANS, J., HARLAND, L., HAUG, K., HERMJAKOB, H., HO SUI, S. J., LAEDERACH, A., LIANG, S., MARSHALL, S., MCGRATH, A., MERRILL, E., REILLY, D., ROUX, M., SHAMU, C. E., SHANG, C. A., STEINBECK, C., TREFETHEN, A., WILLIAMS-JONES, B., WOLSTENCROFT, K., XENARIOS, I. AND HIDE, W. Toward interoperable bioscience data. *In: Nature Genetics*, 2012, Volume 44, p. 121-126.

TENENBAUM, J. D., SANSONE, S-A. and HAENDEL, M. A sea of standards for omics data: sink or swim? *In: Journal of the American Medical Informatics Association,* 2013, Volume 21, no. 2, p. 200-203.

WHALE, A. S., COWEN, S., FOY, C. A. and HUGGETT, J. F. Methods for Applying Accurate Digital PCR Analysis on Low Copy DNA Samples. *PLoS ONE* 8(3): e5177. DOI 10.1371/journal.pone.0058177

Useful websites

www.sbolstandard.org/

www.oxfordjournals.org/our_journals/nar/database/a/

British Standards Institution (BSI)

BSI is the independent national body responsible for preparing British Standards and other standards-related publications, information and services. It presents the UK view on standards in Europe and at the international level.

BSI is incorporated by Royal Charter. British Standards and other standardization products are published by BSI Standards Limited.

Revisions

British Standards and PASs are periodically updated by amendment or revision. Users of British Standards and PASs should make sure that they possess the latest amendments or editions.

It is the constant aim of BSI to improve the quality of our products and services. We would be grateful if anyone finding an inaccuracy or ambiguity while using British Standards would inform the Secretary of the technical committee responsible, the identity of which can be found on the inside front cover. Similarly for PASs, please notify BSI Customer Services.

Tel: +44 (0)20 8996 9001 Fax: +44 (0)20 8996 7001

BSI offers BSI Subscribing Members an individual updating service called PLUS which ensures that subscribers automatically receive the latest editions of British Standards and PASs.

Tel: +44 (0)20 8996 7669 Fax: +44 (0)20 8996 7001
Email: plus@bsigroup.com

Buying standards

You may buy PDF and hard copy versions of standards directly using a credit card from the BSI Shop on the website **www.bsigroup.com/shop**. In addition all orders for BSI, international and foreign standards publications can be addressed to BSI Customer Services.

Tel: +44 (0)20 8996 9001 Fax: +44 (0)20 8996 7001
Email: orders@bsigroup.com

In response to orders for international standards, BSI will supply the British Standard implementation of the relevant international standard, unless otherwise requested.

Information on standards

BSI provides a wide range of information on national, European and international standards through its Knowledge Centre.

Tel: +44 (0)20 8996 7004 Fax: +44 (0)20 8996 7005
Email: knowledgecentre@bsigroup.com

BSI Subscribing Members are kept up to date with standards developments and receive substantial discounts on the purchase price of standards. For details of these and other benefits contact Membership Administration.

Tel: +44 (0)20 8996 7002 Fax: +44 (0)20 8996 7001
Email: membership@bsigroup.com

Information regarding online access to British Standards and PASs via British Standards Online can be found at **http://shop.bsigroup.com/bsol**

Further information about British Standards is available on the BSI website at **www.bsigroup.com/standards**

Copyright

Tel: +44 (0)20 8996 7070
Email: copyright@bsigroup.com

BSI, 389 Chiswick High Road
London W4 4AL
United Kingdom
www.bsigroup.com