

An economy of details: standards and data reusability

Ana Delgado 

TIK Centre for Technology, Innovation and Culture, University of Oslo, Oslo, Norway

*Corresponding author: E-mail: ana.delgado@tik.uio.no

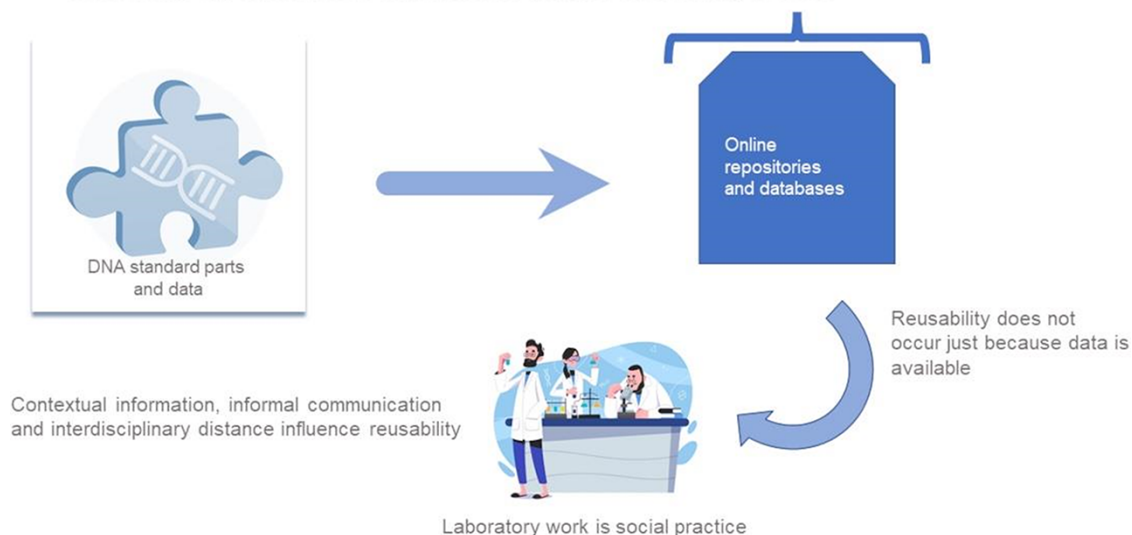
Abstract

Reusability has been a key issue since the origins of the parts-based approach to synthetic biology. Starting with the BioBrick™ standard part, multiple efforts have aimed to make biology more exchangeable. The reusability of parts and other deoxyribonucleic acid-based data has proven over time to be challenging, however. Drawing on a series of qualitative interviews and an international workshop, this article explores the challenges of reusability in real laboratory practice. It shows particular ways that standards are experienced as presenting shortcomings for capturing the kinds of contextual information crucial for scientists to be able to reuse biological parts and data. I argue that researchers in specific laboratories develop a sense of how much circumstantial detail they need to share for others to be able to make sense of their data and possibly reuse it. When choosing particular reporting formats, recharacterizing data to gain closer knowledge or requesting additional information, researchers enact an ‘economy of details’. The farther apart two laboratories are in disciplinary, epistemological, technical and geographical terms, the more detailed information needs to be captured for data to be reusable across contexts. In synthetic biology, disciplinary distance between computing science and engineering researchers and experimentalist biologists is reflected in diverging views on standards: what kind of information should be included to enable reusability, what kind of information can be captured by standards at all and how they may serve to produce and circulate knowledge. I argue that such interdisciplinary tensions lie at the core of difficulties in setting standards in synthetic biology.

Key words: standards; infrastructures; data; reusability; reproducibility; experiment

Graphical Abstract

Which details and information do standards need to include for reusability to occur?



1. Introduction

Reusability has been a key issue since the origins of the parts-based approach to synthetic biology a couple of decades ago. With an ambition to make biology easy to engineer, synthetic biology

from the start directed efforts to standardize biological parts (1). Modular and well-defined parts were expected to enable more predictable design, being context-independent and thus easy to exchange. Synthetic biology was new and fresh; it was envisioned

Submitted: 15 November 2021; Received (in revised form): 12 October 2022; Accepted: 30 November 2022

© The Author(s) 2023. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

to be a different and more open way of doing biology. Open science can be practiced in many ways (2), but the emphasis in this approach to synthetic biology was primarily on enhancing knowledge flow through collaborative and redistributed work to develop a common pool of resources.

The vision of a common repository was pivotal in this, and the BioBrick™ standard part and the Registry of Standard Biological Parts became cornerstones of synthetic biology's early efforts to establish itself as a field (3). The Registry was projected as an infrastructure in which people could deposit their genetic parts and, in turn, use parts that other people had produced (4). Similar to other versions of open science (5), faster knowledge flow was expected to lead to faster innovation. The BioBrick™ became a reference in open-source biology (6), although synthetic biology effectively drew on a combination of proprietary and non-proprietary solutions (7).

The iGEM Competition and the Registry of Standard Biological Parts (today called the iGEM Registry) have functioned as a test bed for synthetic biology as open-source biology. Namely, they have served to try out different versions of assembly standards set by the BioBrick Foundation. Furthermore, iGEM has introduced students to a kind of open-source 'moral economy' of parts production (8, 9). To compete for awards, iGEM teams must contribute to the Registry by depositing new standard parts and/or reusing and improving on already existing ones. Time and many competitions have proven the reusability of parts to be more challenging than initially thought. Lack of parts characterization has been suggested as a reason preventing the reusability of parts (10). In informal discussions, students and other users of the Registry explained to me that it was often difficult to find parts or know what they could be used for and that sometimes iGEM wikis and personal communications were used to access important additional information about standard parts.

Despite these shortcomings, the ideal of the Repository has remained a referent, and over the years a number of local repositories have been launched in different laboratories. Synthetic Biology Open Language (SBOL) probably represents the most comprehensive effort to enhance knowledge flow and parts exchange in synthetic biology. It provides a standard that enables a more complete functional characterization than commonly used formats for deoxyribonucleic acid (DNA) data annotation such as GenBank. Extensive adoption of the SBOL standard remains a work in progress.

Early on, a recurrent issue became clear in the Registry of Standard Biological Parts: that making parts standard and accessible by depositing them in an open repository does not suffice to enable actual parts (re)use. This gap between accessibility and reusability not only pertains to synthetic biology but also applies to other forms of open-source and repository-oriented biology (11). A common problem in all attempts at packaging data in a standard format is that some information is always left out and yet that information may be crucial for a new user. Back in 2008 in a *Nature Biotechnology* paper called 'Setting the Standard for Synthetic Biology', Arkin (12) discussed this problem, focusing on standard datasheets. He noted that:

...there is a science to be developed concerned with the proper packaging and characterization of 'modular' biological activities so that these may be efficiently assemble into applications. [...] Engineers are fond of standards. A good device standard defines sufficient information about discrete parts to allow the design of predictable complex composite systems. [...] Datasheets are an embodiment of such engineering standards.

They contain a formal set of context-dependent, input-output behaviours, tolerances, requirements, physical interconnect 'form factors' (the mechanical requirement for physical incorporation of the device into a system) and other details about a particular part or subsystem. (12, p. 771)

Interestingly, in order to make biological data context-independent, some contextual information needs to be included in annotation formats such as datasheets. Yet standardization always entails some sort of discrimination (13, 14): What information should be included and what information is not needed? When is information sufficient? I argue that there is no possible context-independent answer to that question insofar as what counts as sufficient information depends in every case on the particular setting in which that part is received and meant to be put into use. The question is then: sufficient information for whom? Specific contexts of reception (disciplines, research groups, laboratories, etc.) may require specific sorts of additional information. Both the site where knowledge is produced and the site where it is received are dynamic and distinctive social contexts. In different laboratories, people will surely address different research problems and questions and use different equipment and tools. Given such heterogeneity, the producer of a biological part or other sorts of genetic data may need to knowledge-bit consider the specificity of the receiver's context when packaging it so that makes sense and is reusable in its new context.

This article shows particular ways in which standards are experienced as presenting shortcomings for capturing the kind of contextual information that is crucial to reuse biological parts and data in specific research situations. It draws on a series of qualitative interviews and an international workshop within BioRoboost, a European Commission (EC)-financed project with the specific aim to boost standardization in synthetic biology. This article presents insights into what researchers in laboratories within the BioRoboost consortium actually do in order to use data produced somewhere else and to make their own data reusable. I take a sociological approach, building on Science and Technology Studies (STS). This field has shown that for scientific knowledge to travel, it has historically been 'packaged' in a variety of formats, such as scientific letters, laboratory notebooks and datasheets (15–19). To a large extent, science has been communicated through text-based formats, some of them being more standardized than others yet all of them expressing sets of social conventions.

This article adds to that literature by showing how in laboratory practice, experimentalists may actively calibrate the kind of information new users need in order to bring data back to life and put it in use. Researchers in the laboratory often need more contextual information than the information provided by standard formats. They frequently get that information by recontextualizing the biological part or piece of DNA data in question by validating it, recharacterizing it or rebuilding it from scratch. Such recontextualizations are dependent on experiments and experimental data. In those common instances, data reusability is not independent from experimental reproducibility. Furthermore, when standards do not suffice in providing the information that a particular researcher needs, that person may go on to use informal means—such as phone calls—in order to attain the information. In other words, in real experimental practice, people often rely on nonstandards and informal means to make parts and data reusable.

This article shows how experimental practice includes practical assessments on what information is to be given and what information can be taken as given. In doing those assessments,

researchers project an imaginary user. I characterize that sense of the other and what the other needs in order to make use of some piece of data in terms of an economy of details. The empirical sections of the article suggest that the farther apart two laboratories are in disciplinary, epistemological, technical and geographical terms, the more detailed information needs to be captured by standards for data to be reusable across contexts. The empirical materials suggest that disciplinary distance in synthetic biology between computing science and engineering fellows and experimentalist biologists is reflected in diverging views on what kind of information should be captured to enable reusability, what kind of information can be captured by standards at all and how standards may serve to produce and circulate knowledge. From an open-software perspective, standards appear as facilitating knowledge sharing and flow and so enabling innovation to happen faster. Experimentalists with a background in biology, on the other hand, may value the open-ended character of experiments as core to scientific discovery, interviews show. Furthermore, while for engineers standardizing biology may appear as a practical goal, for experimentalists it may appear as an ideal to pursue but not necessarily something that can be fully achieved in practice. Such interdisciplinary tensions lie at the core of difficulties to set standards in synthetic biology, this article argues.

1.1. Materials and methods

For primary empirical material, this article draws on 15 interviews and the results of a 'Best Practices' workshop carried out within the BioRoboost Project Consortium. BioRoboost was a project financed by the EC (2019–21) aimed at boosting standardization in synthetic biology (20). The consortium included 25 partners from three continents, with researchers mainly based in Europe but also in the USA and Asia. In BioRoboost, I led a work package on the social dimensions of standardization.

The interviews were carried out with principal investigators (PIs), postdoctoral fellows and researchers at different facilities within the consortium, most of whom were synthetic biologists working experimentally. Most of the laboratories and groups within the BioRoboost consortium were working at the intersection of synthetic biology and microbiology; common research areas were metabolic engineering and the development of standard microbial chassis. When interviewees were asked about the data they produced, used or reused, they referred mainly to sequence data, simple genetic constructs and experimental data. In addition, three interviews with SBOL developers recorded during 2015 and 16 were used as supporting empirical material.

The interviews were carried out in two rounds. The first interview protocol included issues broadly related to parts sharing, ownership and standards. After a first round of four interviews, the questionnaire was redesigned to go more in-depth into issues regarding reusability, as it became clear that infrastructural openness and sharing might not be sufficient to result in effective parts exchange. The interviews lasted 20–80 min each. Most of the interviews were carried out during COVID-19 lockdowns and so were conducted virtually instead of the originally planned face-to-face interviews and visits to laboratory facilities. The interview protocol (see [supplementary material](#)) was designed to let interviewees describe what they did during their daily work at the laboratory, with a set of questions that could allow for relatively free talk. This was used as an open-ended guide, with new questions incorporated during the interview if the flow of the conversation required it and questions omitted when they were not relevant and allowance for the interviewees to expand on issues that were more interesting to them. Most interviews were carried

out in English. They were transcribed and analyzed with NVivo qualitative analysis software.

An issue that became quite prominent in the course of the interviews was that researchers (both PIs and others) used means such as email and WhatsApp chats to communicate when standards were not sufficient. To capture this activity, I joined a WhatsApp group in one of the laboratories. In this article, those observations are used as secondary/background material. In addition, I went through emails sent over one of the SBOL developer mailing lists from 2017 to 2018 to get a vivid sense of the kinds of actions taken by these developers to make SBOL more user-friendly, get people to adopt the standard and support data reusability. Finally, in order to gather complementary information on how laboratory researchers use open infrastructures to annotate sequence and experimental data, I asked one PhD student and one postdoctoral fellow in the consortium to show me in an online, recorded demo session how they used Benchling. This is a cloud-based platform for data storage and sharing that seemed to have become quite popular among young researchers in BioRoboost partner laboratories.

To further explore how standards serve to communicate across laboratories and what people do when standards fail to provide the information needed for reusability, I designed, organized and facilitated a workshop. Over 30 PIs and researchers from three continents (Europe, the USA and Asia) attended. The workshop lasted 5 h, was structured in four sessions and drew on a combination of group work and topical discussions. The workshop included a session of a hands-on introduction to Benchling by some of the aforementioned young researchers. The workshop was recorded, transcribed and analyzed. Finally, in order to situate and better understand all these different materials, I drew on my previous research on synthetic biology from 2011 to 2016. This includes multiple conversations with synthetic biologists, visits to laboratory facilities and participation in the iGEM competition as an advisor.

2. Reusability: standards and context

Efforts toward standardization in synthetic biology have attracted the attention of scholars in the field of STS. Studies have produced rich insights into how standardization was considered key to community building within synthetic biology from early on (3, 8). Other works have shown how the adoption of new standards and infrastructures impacts everyday experimental practices and reproducibility (21) and how standards depend upon larger social contexts and forms of life (22). I build on this scholarship to explore how standards may enable the exchange and reusability of biological parts and other types of DNA data and to what extent standards may present shortcomings for those same purposes.

This article looks at standards as enabling communication and the travel of knowledge across contexts. Empirical studies in STS have shown that a condition of possibility for scientific knowledge to travel is that researchers agree on the appropriateness of certain formats to carry it. Since the 17th century, this has mostly consisted of different forms of written text (15, 19). The aforementioned datasheet is one of the formats for packaging knowledge widely used and accepted in today's scientific practice. The scientific paper has long been the most credited carrier, with specific guidelines for how to report experimental results. In today's increasingly data-centric biology (11), open publication of datasets tends to be a formal requirement in scientific publishing, with some standard formats such as GenBank dominating over others. Although publication of data in open repositories

appears as a basic (and sometimes sufficient) requirement for open science, some versions of open science also rely on other formats, such as periodic newsletters in the iconic historic case of the *Drosophila* community (17) or publishing open protocols and laboratory notebooks. The written formats within science follow different degrees of standardization, from open formats such as newsletters—based on widely accepted but unformalized conventions—to more standardized and closed formats such as SBOL in synthetic biology. Each type of format serves to capture information in different ways, with nonstandard formats tending to cover more circumstantial information.

Different formats used to package knowledge entail sets of choices about which information to include as well as how much to constrain or enable individual assessments of how the knowledge will be received, by whom and what information other researchers will need in order to credit the knowledge-bit and eventually use it in the received context. As described by Shapin and Schaffer, for instance, early experimentalists in the 1600s provided an extraordinary amount of detail in their reports (often in private letters), so the peer receptor would have a clear image of the experimental setting in order to possibly reproduce it (23, 24). Detailed reporting served to convey objectivity (23, 24), and it was often supplemented with precise illustrations of the experimental tools used. Scientific textual devices such as letters and journal articles—and today, datasheets—enable and limit the possibilities to provide basic, additional or even circumstantial information wrapped in a knowledge-bit. Arguably, each of those formats enables a certain economy of details that include a choice of specific formats and a situated sense of the amount of detail that an imagined audience or receptor will need to find a particular knowledge-bit appealing, interesting or even useful. This economy of details is performed by particular researchers at particular laboratories working with specific disciplinary traditions, research questions and laboratory tools.

Standards are made to capture and minimize the amount of information needed to transfer a knowledge-bit from one context to another (say, from one laboratory or research group to another). Ideal standards should be economic, reducing the amount of circumstantial detail, on the premise that circumstantial detail can produce noise. Yet, different 'ways of knowing' (25), such as experimental biology and engineering, may perform different economies of detail. In practice, the extent to which knowledge travels by means of a standard depends on factors that exceed the standard alone and may vary from case to case.

Furthermore, collaborations across disciplines may require substantial circumstantial detail for a knowledge-bit to make sense far from the context in which it was produced. STS scholars Florence Millerand and Geoffrey Bowker analyzed the challenges of interdisciplinary data exchange in environmental sciences. They looked at infrastructures and standards developed to enable exchange within those disciplines and concluded that:

A central problem here is that the storage, access to and evaluation of the validity of data are extremely dependent on the ways in which the data has been collected, labeled and stored. While it may be possible for two colleagues in a discipline to share information about their data with a simple longhand note, there is unquestionably a need for more documentation in the case of pluridisciplinary teams working over multiple sites and scales. (26, p. 150)

This suggests that the farther the receptor, the more circumstantial detail of the context of production may be needed in

order to make sense and possibly make use of a bit of data. Such distance between contexts of reception can be geographical, epistemological, technical or disciplinary, the empirical sections of this article show.

Looking at empirical materials from synthetic biology, the following section of this article explores how a certain version of an economy of details is performed in biological experimental practice. This article attends to what people do when standards are not enough to capture the information needed for knowledge-bits to travel and be reused across contexts, be they standard biological parts or other sorts of DNA data. Interdisciplinary frictions arise not only in the form of disagreements about how much detail is needed for knowledge to travel but also in relation to what standards in synthetic biology can actually capture.

2.1 Capturing context through standards: dealing with emerging biological systems

As in any other life sciences field, laboratory research in synthetic biology is largely dependent on standards of different sorts. Discussions during several BioRoboost consortium meetings made clear how partners generally agreed on the importance of improving standardization in synthetic biology. They often diverged, however, on what exactly was to be standardized (metrics, assembly methods, etc.), pointing to a variety of needs and priorities within their different laboratories. Divergent opinions were also noticeable regarding the extent to which standards were seen as doing a good enough job in capturing the biological information needed in practice for actual data reusability, interview material indicates. Arguably, such divergence points to interdisciplinary frictions in the field (27). While synthetic biology developed from an ambition to turn biology into a true engineering discipline, this emphasis on engineering has been somewhat contentious and challenged by practitioners with backgrounds in the life sciences. The extent to which biology can be turned into standard modular parts that can work in a context-independent fashion—as in software engineering—has been a matter of debate (28, 29). Moreover, while discrete modular parts may be easier to capture through text-based standards such as data sheets, emerging and context-dependent biological systems make such standardizing endeavors more challenging, the empirical materials of this article suggest.

When interviews included questions on the functioning of biological systems, laboratory leaders tended to appeal to intrinsic features and mechanisms of biological systems, such as flexibility, adaptability, evolution and emergence, referring in one way or another to life's capacity to change over time. In that line of argument, the project to turn biology into standard modular parts appears as a guiding ideal, a place to direct efforts rather than a fully achievable practical goal. Interview participants described different ways in which they worked toward standardization and how biological contexts made standardization difficult. Biological contexts were described in multiple ways in the interviews. For example:

So, I think that... the step in bringing standards to biology is to focus in metrology, and in this area of engineering that has to do with how to measure things, how to measure functions, how to measure properties. So, even if we agree on a standard to measure things in biological systems, to me that will be a fantastic achievement. But if you go further than that, then you have to go back and rethink a few questions, because then things become very complicated or very intricate. I mean, not impossible to tackle, but certainly it's not straightforward. And also in

biology you have this phenomenon that we call emergence. So, that means that two plus two is not always four. It can be less than four, it can be more than four, and that's something that engineers cannot live with...That's the basis of biology. (Spain, PI, BioRoboost 15)

In this quote, emergence is pointed out as one of the features that make biological systems difficult to standardize, and yet standardization remains a desirable goal. The PI in the interview quoted above emphasized the importance of standardizing metrics to make data comparison and exchange across laboratories easier. However, standardization in metrics is in practice challenging as different laboratories frequently pursue different research questions and may need to measure different things. As another PI explained, different laboratories may have different equipment, some may be more able than others to update their equipment and in that case some new measurements will perhaps be required. These arguments bring together limitations imposed by biology with challenges imposed by experimental tools and conditions. Differing biological contexts and technical equipment challenge the reusability of standard parts and DNA-based constructs. The next section provides insights on the kind sociotechnical challenges experienced by experimentalists in that regard. Concerning biological systems, and the extent to which they can be captured through standards, interviews show variation from this being a problem that can be somehow 'tackled' to it being an unresolvable problem. A PI in Germany, for instance, expressed doubts that biology could ever be fully standardized, precisely because of the holistic nature of biological systems. This PI specifically referred to evolutionary contexts as troubling standardization:

That is all directly related to the way we do science. As much as I understand standardization as a concept, let's say from physical and engineering disciplines [...] we should not forget that biology is a little bit different. In physical terms, a transistor is a transistor, right? It will never evolve into something new. But in biology, a transistor is something that has the potential to evolve into something new. It is only as accurate as it needs to be. There is always this flexibility built into it to do something new. [...] So typically proteins, but also actually many other things like genes, are evolved to be as efficient and as correct or specific as it needs to be. But not more. Because otherwise you sacrifice the potential to evolve. [...] We take this as an essential part of a biological part: that it has flexibility built in, due to the fact it needs to evolve further. This is one of the problems I see with you, it is meta-data annotation. It is how can you assess the full potential of a biological part, because you focus on the main function of the main part, but you very often do not focus on the other functionality. And those are pretty hard to grasp. Because every part has a different evolutionary history. It comes with a different context. (Germany, PI, Interview 12)

In the previous two quotes, interviewees established a clear disciplinary distance between biology and engineering. Nevertheless, interviews point to variation among biologists too, with some emphasizing standardization of biological systems as an ideal to pursue. Interviewee 12 emphasized that biological systems are fundamentally context-dependent, dynamic and ultimately difficult to predict, which also troubles standardization in data annotation. This open-ended character of biological seems to match the rather flexible and open way of knowing in the

experimental discipline of biology (and in this case, specifically microbiology):

I am given a lot of scientific freedom, and I do not think we should be too restricted by being too standardized. We are not industry; we are still discovery-driven. And there might be things which cannot report easily, right? The kind of science we're doing is very much explorative. It is very much discovery-driven. If you do not know how big a field this is, you cannot draw a map. If you don't know there is a new force of nature, you cannot describe it. I think you should keep this flexibility, but I think for understanding and comparing data, it is helpful. But over-standardization might be a little bit problematic. (Germany, PI, Interview 12)

In the same line of thought, another PI in a different country explained that he knew PhD students and postdocs in his laboratory might be doing some other projects 'on the side', but he allowed and even encouraged it because that was research out of genuine curiosity. Freedom and relaxed standardization appeared to some as a productive resource in science. Some researchers explained that they would not request the researchers working in their groups to shift to digital laboratory notebooks or report their experiments in standard formats for similar reasons, because something would be lost:

If everything in the lab was standardized, there would be also something to lose because some great discoveries and surprises have been made because of unwanted variations. That someone did something different by accident or because he didn't know and then you get an unusual result, you follow up and it turns out that you have discovered something really exciting [...] Robots could not know that that was exciting. [...] You need a person to understand that that was exciting. (Germany, PI, Interview 11)

The quotes above point to the context-dependent nature of biological systems to explain variation. Interviewee 11 pointed to the need to have a person (versus automation/standardization) to see potential in unexpected variation arising in experimental practice. 'Unwanted variations' appear as something positive because they are a main source of scientific discovery. This view is in line with Rheinberger's account of the history of molecular biology in which experimental systems are characterized as information generating systems, in which 'differential reproduction' (30) is key to the occurrence of variation. In Rheinberger's words, 'All innovation, in the end and in a very basic sense, is the result of such reproduction. Reproduction, far from being simply a matter of securing appropriate and reproducible boundary conditions for the experiment, characterizes scientific activity as a material process of generating, transmitting, accumulating and changing information' (30). Resonating with this, most of the researchers I talked to, especially those with a training in experimental practice within the life sciences, supported a view of experiments as open-ended systems. This can be seen as contrasting with open-source ideals in which standards enhance knowledge flow and innovation. Furthermore, the open-ended nature of both biological systems and experimental practice challenge attempts to transcribe biological entities into standard formats, be they simple datasheets or more complex representational formats such as SBOL. As the next section further explores, this is something that experimentalists often experienced in their everyday laboratory practice.

2.2 Capturing context beyond standards: reporting, annotating, phone calls and emails

When interviewees working in laboratory facilities were asked to describe what kind of data they worked with, they referred mainly to sequence data and genetic parts and constructs. When talking about what they did in order to produce or reuse data, they often referred to experiments, with experimental data mainly captured in laboratory notebook descriptions and protocols. When talking about how they produced data, it was difficult to draw a clear line between experimental data and other kinds such as annotated sequence data. If they were to work with sequence data annotated in a simple format such as GenBank or a standard biological part such as a BioBrick™ produced by another research group, they would commonly search for or request additional information that would frequently also include experimental data. They reported that when they used data produced in other laboratories, they would often test those data, validate it or recharacterize it before reusing it. They needed to bring those data back to life in experiments and re-make it to some extent in order to adapt it to the kind of research questions and laboratory equipment they worked with:

I would also prefer to use other people's data, rather than have to regenerate it ourselves. But I am struggling to think of times where I would do that. Normally if we want to use someone else's data, then we are going to generate our own data with it. When I am interested in someone else's data, for my kind of project, I want to look at their data to understand that if I get the DNA part, the genetic parts, or all the strains from them, will that improve the system I am trying to build? So I'll look at their data and evaluate what they have done, based on their data. And then request the stuff, and then probably the first thing we'll do is repeat the experiment to see if in our hands, under our conditions, in the context of what we want to do, does it still perform as well. (UK, PI, Interview 9)

Statements such as this one on how the interviewees produce and reuse data indicate that it is difficult to say in practice where experimental data end and where sequence data start. In other words, there is no clear line between data reusability and experiment reproducibility: actual reusability in practice often depends on putting experimental data into use. Perhaps, this is more so for synthetic biology than for fields that are more markedly data-driven, such as systems biology. This raises the question: to what extent and in what sense is synthetic biology a big data science and to what extent is it experiment driven? Despite being inspired by the open-software movement, despite the iconic place that repositories have played in creating the field and despite aspirations at de-coupling design and construction, laboratory experimental practice seems to remain core to knowledge production in synthetic biology. When trying to reuse data produced somewhere else, researchers would often reproduce experiments and request additional information. In turn, when asked what they did in order to make their data reusable, researchers referred to the experience-based knowledge of the experimentalist, which includes knowing how to annotate and report experiments. In the interviews, the idea that knowledge could be fully captured in standards was repeatedly challenged:

There was a publication a few years ago where they had done a factorial analysis of what are the important factors in terms of microbial growth. If we could capture the information for any given experiment, or relating to those important factors, then

you would have a much better understanding how the experiment was done, in a way which was important for reproducing it. And at the moment, you look at the materials and methods in the papers and basically it is: overnight culture diluted to one in a hundred, taken to a plate, shaken, and grown in the plate reader. That is about it. It is very sparse. (UK, PI, Interview 4)

The PI who gave that quote was based in the UK. Another PI based in Spain also referred to notebook annotations of the type such as 'overnight culture'. He explained that 'overnight' has a very different meaning in Spain than in other places, as in Spain people tend to work longer hours. Effectively, 'overnight' is often shorter in Spain. This example shows one way in which experiments are situated. All the participants interviewed agreed that annotations in notebooks should be detailed and precise in order to capture the specificities of each experimental situation. Most PIs saw advantages in digitalizing and standardizing notebooks; at the same time, they emphasized individual researchers' freedom to decide what suited them best and to develop their own ways when it came to annotating experiments:

Science is a creative discipline. You need to encourage people to do things in the way they want to do them, with their own freedom. I have purposely never set one specific one way to do a lab book with my group. [...] They can use online or paper lab books, whichever they would be taught from their previous education. Different places teach different things. (UK, PI, Interview 9)

I interviewed researchers who preferred paper notebooks and gave reasons for that. One of them explained that when he goes back through the pages of his notebook, he 'remembers' things that went wrong. The way he described it, he recalled the situation of the experiment as a kind of embodied memory. I also talked to some younger researchers (PhD students) who were in favor of digital notebooks. I conducted two demo video-call sessions with them, asking them to show me how they used digital notebooks on the Benchling platform. They explained the advantage of digital notebooks as being much more structured, enabling collaboration and providing an overview of all that is important for a project, as all parts are linked and available in the same platform (pictures of gels and PCRs, sequences, protocols, scholarly references, etc.).

Still, most people in the interviews and talks agreed that annotations in notebooks were often insufficient. Lack of complete annotations has to do not only with a lack of precision in definitions, as we see in the quote previously (UK, PI, Interview 9), but also with difficulties in capturing all the important details that are key to reproducing experiments and thus for researchers to recontextualize and reuse data in a different laboratory. The amount of detail in annotation would vary depending on the epistemic distance between laboratories, the tools at use and the 'area' of interest:

We try to define a sort of minimal sets of additional information that needs to be associated to the data. This is different if you want to share in your lab, if you want to share in your specific community, so with colleagues working in the same area with the same organism. They have much the same experience; they need perhaps less additional information to make sense of it [...] I could imagine that for some model organisms and for some broader systems-biology approaches or for some very basic tools of model organisms, a community can agree on making a sort of big data sets under exact conditions. Perhaps in a centralised place to have this model organism, the larger

dataset would make it available. As soon as you start to spread it in the community and you have different hands, different labs, different media, different glass, different dishwashers for your labware, it could already make a difference [...] Okay, now I'm in a different lab. I get this very complete description of how this organism was grown and I want to repeat it. To reproduce something and use it for my own work. And then I may realize it doesn't grow as they have described it. But you had a very complete description set. I did everything the same. I used the same medium. I used the same. And when it comes to the level of detail and what level of detail is required...sometimes there are really, really tiny differences. [...] (Germany, PI, Interview 11)

After explaining how small divergence in following a protocol may make a big difference, the same PI went on to explain how some researchers have the capacity to know what is the 'right level of detail' they need to note down, they can differentiate between what is and is not important to report. She described this as part of being an 'experienced' experimentalist. She argued that discriminating between what is important information and what is not is something that machines will never be able to do. This and one other PI in a different laboratory described such intimate way of knowing experiments not really as a knowledge but as a 'feeling':

Yes, but you have people who are less or more careful when doing these things and recording it. But it is also that a very high level of details is not the best to do. It is to find exactly the right level of detail. And this depends also very much on the experience of the person. If you have worked for 30 years in the lab, you have a sort of feeling – and I really say feeling because it is not really knowledge, you have a feeling – You need to understand the process and what steps really matter. [...] Where it makes a difference if your tube has a very thin wall or thicker wall. Because it makes a difference how quickly in a specific PCR machine you have a heating up, and that has an influence on the type of library you get afterwards. So in our protocols, it is a critical step where we say what type of tube, from which vendor, how many segments, which PCR machine, which block of the PCR machine and the age of the block, because the [...] elements get older over the time. We record this and it is very important [...] but if we would do this [to record everything] for every step, it would be useless and it would confuse people, because they would not know anymore what is an important step and where we really need to follow the protocol very very exactly, and other steps where if someone calls you, you can answer the call, it doesn't matter. As long as we don't have robots in the lab – and by robot, I do not mean the lab automation we are using, I mean a robot doing everything [...] As long as you have this (people), you will have variation, because experimentation makes it. You can't do everything always exactly the same way. In some protocol steps it makes a difference, in others not. A person with lots of experience has a feeling for that, and an understanding. And a beginner in the lab has no idea at all about that. (Germany, PI, Interview 11)

This researcher explained how 'a feeling for the experiment' was something that one can learn. In the interview, this 'feeling' was described as a sort of intimate knowledge of experimental situations in all they entail, including knowledge of particular laboratory tools and equipment, how sensitive they are and how they behave and age. This kind of intimate knowledge of the materials and temporalities of experiments has been reported before as a kind of knowledge that can be only acquired through practice

(31). The experimental situation was described by this and other PIs as very rich and full of details. However, part of the knowledge of the experimentalist is to know how to distinguish between details that are important to be able to reproduce the experiment and those that are not so important. Experienced experimentalists have the capacity to discriminate, deal with variation and annotate just what is important. They can make decisions such as (i) which steps are routinary and easy to repeat, which can be modulated and adjusted and at which small mistakes are likely to make a difference; (ii) how to record this so that the person receiving the message can get a sense of the differences that matter and (iii) developing an imaginary of the receiver, including an assessment of what they should already know. As I was told, people working in the same laboratory with the same problems and tools need less detailed annotations. The researcher who gave the above-mentioned quote (Germany, PI, Interview 11) explained that the farther the distance between two laboratories—in terms of the kind of equipment they use, the kind of research questions, etc.—the more detailed annotations are needed for people to reuse data and reproduce experiments. As mentioned previously, this has been noticed by STS colleagues for fields beyond synthetic biology (27); this problem of incomplete annotation is a recurrent one in data science.

I suggest considering this in terms of an 'economy of details', which relies on a sense of how much information to give and how much to take as given. This economy of details depends on the experimenter's close knowledge of the experimental situation and a whole laboratory machinery: laboratory tools and metrics, recording and displaying devices (such as pictures of gels and PCRs), registry forms and specific formats (such as Excel forms) and specific annotation platforms (paper notebooks or online platforms such as Benchling).

When annotations remain incomplete, data users may need some additional information in order to be able to effectively use the data. In such a case, interviewees reported, they would simply make a phone call or, more commonly, send an email. Researchers often requested data from other laboratories instead of retrieving it from open repositories. When requesting data from other research groups, particularly sequence data, it was common to request additional information on how the data were produced. Interestingly, for fairly standard and widely used formats such as Excel sheets to be functional, they often depend on informal communication channels, including face-to-face encounters. It was also through personal communications that new students in the laboratory trained themselves as experimentalists and learned to discriminate which details matter:

You can try to make a protocol, but it is not always easy to tell that. Sometimes it is also hard to know what you have to tell. So if you make an SOP, standard operation procedure, of how you have to grow cells or how you have to transform them. You can put all the elements that you know, or that you think is logical, but sometimes you do something that you don't realize but that is important within the SOP. Most of the time it is these little things that are not obvious that you can only transfer or communicate when you do it together with someone. Then they can see it. It is very hard to put that on paper, because you don't realise it, it is only because they mimic you [...] They take over that same behaviour. I don't know if I am explaining it well, but it is sometimes really small things you do. Putting something on ice or not. Or, it is on ice, but before, you flip it, and it isn't really described in the SOP but maybe that flipping is important. That is something that you see when you

are communicating informally by showing it. That you shadow someone that is explaining it. It is always easier, the transfer of information is always easier. (Belgium, PI, Interview 13)

This researcher described how students become experimentalists, i.e. they learn all the details that matter, by physically imitating ('shadowing') someone else. In the STS literature, this kind of noncoded know-how is characterized as 'tacit knowledge' (32). With its emphasis on its partial and embodied features, it can also be described as 'situated knowledge' (33). It is beyond the scope of this article to enter into a discussion on tacit knowledge *per se*. Rather, what interests me with regard to reusability is to explore what people actually do in attempts to 'capture' what is missing to be able to use the data in other contexts and how standards may often not be enough for that purpose. As the interviews reveal, when information is not encoded, it is missing. When it is difficult for people to interpret notebooks, protocols or sequence data, they find ways to figure that information out, typically by getting in touch with the person who produced the data. Yet on other occasions, as another researcher remarked, you simply cannot know what is missing 'because you don't know that it matters. And nobody in the world could have known that it matters'. Not knowing what you do not know, uncertainty or indeterminacy may appear as non-desirable features of experimental situations designed (by definition) to take place under controlled conditions. However, experimentalists may also value the open-ended character of experiments, which can even appear as productive and key to curiosity-driven science, as seen at the end of the previous section. Dealing with uncertainty and variation, together with the capacity to discriminate between relevant and less relevant details, is part of the experimental practice and the economy of details underlying scientific reporting and annotations.

3. Recapturing: 'things that are missing'

It would be inaccurate to present a picture of synthetic biology in which people relate to data reusability exclusively in the ways described previously. What the empirical materials of this article show is that although researchers in the field may generally acknowledge the importance of standards for knowledge sharing, there are divergent views regarding the extent to which standards serve to capture some types of information necessary for actual reusability. Disciplinary background and age may be influencing such variation. For instance, in an interview with a relatively young associate professor with a background in engineering, he explained how in his group they are using standard electronic notebooks in Benchling format. He pointed out, 'That is why we have this big problem of reusability in biology, people forget' (UK, PI, Interview 10). Forgetting to note down some important piece of information is a common way in which a researcher would fail to make her data transferable, and standard formats can help to avoid that problem, this interview indicated. Further research on the use of platforms, such as Benchling.com and SnapGene.com within BioRoboost consortium laboratories, suggested that junior experimentalists may be more interested in using those platforms than researchers with more seniority. Three PhD students and one postdoc working in different laboratories in the consortium were invited to present Benchling during the workshop. They claimed that the main advantage of Benchling was in providing easy access to structured information. Information could be structured as 'whole projects' where you could link and trace different parts. As one user explained, 'You can link protocols into the notebook so you do not need to write something twice, you can add schemes,

copy pictures. You can also use the lab notebook to draft papers (Section 1.1). You can share these notebooks with students and colleagues and export files and folders' (PhD students, Belgium, User 2). The presenters explained the main advantages of a platform such as Benchling.com as being the ability to 'link' sequence data, annotation, laboratory notebooks and protocols, as well as the possibility to go back to all the records. In addition to enabling an augmented sense of access, Benchling standards can be experienced as means to aid memory and to avoid unnecessary detail, for the sake of enhancing communication.

Moreover, the kind of embodied and situated knowledge emphasized by senior experimentalists with a primary background in the life sciences (previous section) may appear in an engineering view as a source of disturbance rather than something positive. It may even be overrated, this quote from a software engineer participating in the workshop suggests:

There is almost mysticism about implicit knowledge. That in many cases is actually (exacerbating?) over a fragility in the practices of a laboratory. [...] It is often less about this magic of the implicit knowledge and more that there is a problem hiding that you didn't see until you try to reproduce it somewhere else. [...] Bringing in process control and quality controls are something that is not used very much. As soon as you start using them, you start exposing things that are missing in our typical laboratory practices, not because people are bad or sloppy but because we are dealing with extremely complex things and unless you have a machine or other type of checking system that are checking all the things that are easy to miss [...] Because there is so many ways to make a mistake or to introduce a hidden fragility into your experiment that just eliminating those goes a huge way toward making things reusable and reproducible. (US, Senior Researcher, Workshop Participant 21)

This workshop participant argued that the 'implicit' knowledge (as he called it) that might be making reusability and reproducibility difficult could be made explicit by applying quality checks along the experimental workflow. While implicit knowledge may be more an obstacle than an advantage, automation would help such quality checks. In this view, standards can capture all the necessary information for data reusability to be successful. During the workshop, another participant replied to this comment by reporting that they used workflow quality checks in her company and experienced the adoption of standard quality checks as something very positive. Here, we see a contrast with the interviews, in which the difficulties involved in capturing contextual knowledge were recurrently emphasized.

4. An economy of details: reusability, standardization and openness in science

Drawing on a selection of empirical materials, this article has provided insight into how a standard may have shortcomings for the purpose of capturing sufficient information to enable the reuse of biological parts and other DNA-based data. This article has shown what researchers do in such cases: from reproducing experiments to get a closer knowledge of the data and how it may work in context to actively requesting additional information regarding the circumstances in which the data were produced. All of those attempts at recontextualizing data are attempts to gain circumstantial knowledge. Standard formats such as datasheets are made to be economic in facilitating knowledge transfer. Yet in

daily experimental work, such formats may not be sufficient and researchers may need to do context-recovering through, for instance, informal communication. The empirical findings in this article suggest that the role and importance of informal communication channels and other recontextualization practices in daily laboratory work may have been underestimated in synthetic biology debates on standardization. I have characterized attempts at capturing context in terms of an ‘economy of details’.

As shown, effective data reusability may rely on having access to relevant circumstantial detail. Particular text formats that are used to facilitate knowledge travel, such as datasheets, enable a certain economy of detail. They facilitate the work of data producers, but they may also constrain the amount of detail that can be conveyed. When choosing particular formats, recharacterizing data to gain closer knowledge, annotating data, reporting on experiments or asking for additional information, researchers enact an economy of details. This points to a multiplicity of ways in which researchers practice reusability in their laboratories. Such diversity also reflects differences in how people consider biological, social and technical contexts as being present in laboratory practice and difficult to capture with standards. Experimentalists and software engineers may certainly differ in ambitions to encapsulate contextual information (be it, biological or sociotechnical contexts) and to make it transferable. As defined here, an ‘economy of detail’ expresses a sense of how much and which contextual information is important for an imagined user. As the empirical materials suggest, that sense may change from discipline to discipline, from experimentalist practice to software engineering. For the experimentalists I interviewed, data standardization may appear as an ideal to pursue, while for the software engineers it may appear as a practical goal. This also reflects differing economies of detail with differing approaches to how much and which information needs to be given and how much and which can be taken as given in data exchanges for data to be reusable in practice. As we have seen, a ‘feeling for the experiment’, intimate knowledge of tools and equipment and a sense of how to go about experimental variation (including how to report it) are parts of the experimentalist’s economy of details. Minimizing the amount of circumstantial detail on the behavior of living systems under experimental conditions or on how the experiment was conducted may appear as constraining from the point of view of the experimentalist with a biology background. From a software engineering point of view, too much circumstantial detail may more easily appear as noise. The empirical materials of this article point to disciplinary disagreements as to how standards may serve to improve knowledge flows between research facilities. From an open-source perspective, standardization accelerates knowledge flows and data reusability, leading to faster innovation. Experimentalists in biology, as we saw, may be quite closely attached to their experiments and may see them (particularly in their open-ended character) as a source of scientific novelty.

The insights that this article provides are in line with previously identified interdisciplinary frictions (27) and ‘cultural divide’ characterizing synthetic biology as a field (34). Two ways of organizing scientific collaboration are at play in these interdisciplinary relations. On the one hand, open-source science organizes around redistributed work, sharing and rapid exchange. Experimental work, on the other hand, builds on intimate knowledge of biological entities and tools, runs on its own tempo and has its own way of organizing knowledge exchange that depends on a different economy of details. The kind of disciplinary tensions identified here is key to the problem of setting standards in synthetic biology and deserve further investigation.

Supplementary data

Supplementary data are available at SYN BIO online.

Data availability

Not applicable.

Material availability statement

Not applicable.

Funding

European Union through the BioRoboost Project [H2020-NMBP-TR-IND-2018-2020/BIOTEC-01-2018 (CSA), Project ID 820699].

Acknowledgments

Many thanks to the BioRoboost consortium members and especially to those who dedicated time to participate in the interviews and workshop. Thanks to Thomas Måløy and to the anonymous reviewers for their valuable comments on previous versions of this article.

Conflict of interest statement. No potential conflict of interest was reported by the author.

References

1. Endy, D. (2008) Can we make biology easy to engineer? Feature commentary. *Ind. Biotechnol.*, **4**, 340–351.
2. Levin, N. and Leonelli, S. (2017) How does one “open” science? Questions of value in biological research. *Sci. Technol. Human Values*, **42**, 280–305.
3. Campos, L. (2012) The BioBrick™ road. *BioSocieties*, **7**, 115–139.
4. Torrance, A. W. (2017) Better to give than to receive: an uncommon common in synthetic biology. In: Katherine J, Strandburg B, Frischmann M, Madison MJ (eds). *Governing Medical Knowledge Commons*. Cambridge University Press, Cambridge, UK, pp. 192–222.
5. Hope, J. (2008) *Biobazaar. The Open-Source Revolution and Biotechnology*. Harvard University Press, Cambridge Mass.
6. Hilgartner, S. (2012) Novel constitutions? New regimes of openness in synthetic biology. *BioSocieties*, **7**, 188–207.
7. Calvert, J. (2012) Ownership and sharing in synthetic biology: a ‘diverse ecology’ of the open and the proprietary? *BioSocieties*, **7**, 169–187.
8. Frow, E. and Calvert, J. (2013) ‘Can simple biological systems be built from standardized interchangeable parts?’ Negotiating biology and engineering in a synthetic biology competition. *Eng. Stud.*, **5**, 42–58.
9. Delgado, A. (2016) Assembling desires: synthetic biology and the wish to act at a distant time. *Environ. Plan. D.*, **34**, 914–934.
10. Vilanova, C. and Porcar, M. (2014) iGEM 2.0—refoundations for engineering biology. *Nat. Biotechnol.*, **32**, 420–424.
11. Leonelli, S. (2016) *Data-Centric Biology: A Philosophical Study*. The University of Chicago Press, Chicago.
12. Arkin, A. (2008) Setting the standard in synthetic biology. *Nat. Biotechnol.*, **26**, 771–774.
13. Higgins, V. and Larner, W. (2010) *Calculating the Social, Standards and the Reconfiguration of Governing*. Palgrave MacMillan, NY.
14. Lampland, M. and Leigh Star, S. (2009) *Standards and Their Stories: How Quantifying, Classifying and Formalizing Practices Shape Everyday Life*. Cornell University Press, Ithaca and London.

15. Latour, B. and Woolgar, S. (1979) *Laboratory Life: The Social Construction of Scientific Facts*. Sage Publications, Beverly Hills.
16. Kelty, C.M. (2012) This is not an article: model organism newsletters and the question of 'open science'. *Biosocieties*, **7**, 140–168.
17. Kohler, R.E. (1994) *Lords of the Fly: Drosophila Genetics and the Experimental Life*. University of Chicago Press, Chicago.
18. Asdal, K. and Cointe, B. (2022) Writing good economics: how texts 'on the move' perform the lab and discipline of experimental economics. *Soc. Stud. Sci.*, **52**, 376–398.
19. Shapin, S. and Schaffer, S. (1984) *Leviathan and the Air-pump: Hobbes, Boyle, and the Experimental Life*. Princeton University Press, Princeton, NJ.
20. BioRoboost. <https://standardsinsynbio.eu/> (28 June 2022, date last accessed).
21. Meckin, R. (2020) Changing infrastructural practices: routine and reproducibility in automated interdisciplinary bioscience. *Sci. Technol. Human Values*, **45**, 1220–1241.
22. Mackenzie, A., Waterton, C., Ellis, R. et al. (2013) Classifying, constructing, and identifying life: standards as transformations of "the biological". *Sci. Technol. Human Values*, **38**, 701–722.
23. Shapin, S. (1984) Pump and circumstance: Robert Boyle's literary technology. *Soc. Stud. Sci.*, **14**, 481–520.
24. Schaffer, S. (1998) The leviathan of Parsonstown: Literary technology and scientific representation. In: Lenoir T (ed). *Inscribing Science: Scientific Text and the Materiality of Communication*. Stanford University Press, pp. 182–223.
25. Pickstone, J.V. (2000) *Ways of Knowing: A New History of Science, Technology, and Medicine*. University of Chicago Press, Chicago.
26. Millerand, F. and Bowker, G.C. (2009) Metadata standards. Trajectories and enactment in the life of an ontology. In: Leigh Star S, Lampland M (eds). *Standards and Their Stories*, Cornell University Press, Ithaca, pp. 149–165.
27. Delgado, A. and Porcar, M. (2013) Designing de novo. Interdisciplinary debates in synthetic biology. *J. Syst. Synth. Biol.*, **7**, 41–50.
28. De Lorenzo, V. (2011) Beware of metaphors: chasses and orthogonality in synthetic biology. *Bioeng. Bugs*, **2**, 3–7.
29. Porcar, M. and Peretó, J. (2016) Nature versus design: synthetic biology or how to build a biological non-machine. *Integr. Biol.*, **8**, 451–455.
30. Rheinberger, H. (1997) *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*. Stanford Univ. Press, California, pp. 75–76.
31. Fox Keller, E. (1983) *A Feeling for the Organism. The Life and Work of Barbara McClintock*. W.H. Freeman and Company, NY.
32. Collins, H.M. (1974) The TEA set: tacit knowledge and scientific networks. *Sci. Stud. (St Bonaventure)*, **4**, 165–185.
33. Haraway, D. (1991) *Simians, Cyborgs, and Women: The Reinvention of Nature*. Routledge, NY.
34. Nelson, B. (2014) Synthetic biology: cultural divide. *Nature*, **509**, 152–154.