Lost in Translation: Can We Talk About Big Data Fairly?

Matilde Fontanin(✉) and Paola Castellucci

University La Sapienza, Rome, Italy
{matilde.fontanin,paola.castellucci}@uniromal.it

Abstract. Big data and data science are global, there is no alternative in our connected, digital world. Yet, for a truly open and fair science, cultural biases and different opportunities across different countries must be taken into consideration.

English has become the international language for the scientific debate: a single language is most convenient, moreover it is undergoing a process of refinement and adaptation to the science register. On the other hand, laboratories are populated by researchers from all over the world, and much research takes place in non-English-speaking countries, where research tradition often develops moving from different perspectives, influenced by the cultural context.

A fair and open science would miss an opportunity if it did not take into consideration the multilingualism and multiculturalism of the researchers as individuals and members of specific communities, and could also waste precious time and energies, as language barriers prevent cooperation.

The paper will discuss the above-mentioned issues with examples and reflect on the changing role of librarians and information specialists within a global scientific community.

Keywords: Communication of data science in English · Multilingualism · Decolonising digital libraries

1 Overview

Big data and data science rely on a large international scientific community; scientists from different disciplinary fields, of multilingual and multicultural backgrounds share, openly and fairly, data including "*original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material*" as in the Berlin Declaration [1].

In cyberspace, language is both a source of data and a tool to discuss them globally, it is necessary to put them in context, reason about them and to create metadata and, since researchers from different countries speak different languages, the need for a common language is as strong as the one for a shared technical jargon. Even on common disciplinary grounds, cultural differences remain, related to gender, provenance, social and educational backgrounds: to work together effectively, a linguistic and cultural mediation is needed.

M. Ceci et al. (Eds.): IRCDL 2020, CCIS 1177, pp. 35–46, 2020.
https://doi.org/10.1007/978-3-030-39905-4_5

Last year at IRCDL the authors [2] had reasoned on the barriers to a truly open access to science and the role of librarians. We will focus here on the predominance of English as a language for international communication among scholars, though if "*Our mission of disseminating knowledge is only half complete if the information is not made widely and readily available to society*." [1], the concept of *cultural justice*, as defined by Ross [3], connected to *citizen science* and cultural context is in the background – though it would deserve a treatment of its own, which is beyond our scope. An ethical approach is needed if we share the view that "*Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits*" [4]. There is agreement that "*the Internet now offers the chance to constitute a global and interactive representation of human knowledge* [1]: in scientific communication, open to scholars and laymen, this implies different views, languages and jargons, as well as resorting to different codes, as images and even videogames [5]. Communicating to the general public research that may impact on their lives entails choosing carefully the tools; moreover, being aware of cultural and language barriers is beneficial to the multicultural community of scientists too. In this respect, librarians and other information workers feel involved since, according to IFLA Code of Ethics [6], their mission is to organize information and the services to access it for their users – be they scholars or laymen.

The paper will discuss with examples these issues, reflecting on the role librarians can play in a global scientific community and closing with some proposals.

2 Language Issues in the International Science Community

2.1 Discourse About Science: Language and Political Issues

The data collectively produced and analyzed are expressed in many formats, not necessarily text. Real objects and visual formats may be self-evident, nevertheless, to understand the deeper meaning of objects or the value they represent, a narration is necessary, as Hans Rosling demonstrates, creatively unleashing the power of statistical data [7] to popularize science. Data mapping the *geopolitics of science* show a definite, political hegemony of English. The ensuing marginalization and scarce visibility of other languages and cultures could be tackled effectively if we considered *openness* in the wider, political sense described by Chan [8], one of the signatories of the BOAI statement in Budapest [9].

Machine learning and Artificial Intelligence help analyze and understand the digital universe of data that people produce everyday simply by living online. The span of these fields of studies is increasingly multidisciplinary. A new, global science cannot ignore that AI affects the way people build their selves in a digital world. As Cheney Lipold suggests, "*Who we are in terms of data depends on how our data is spoken for. And in general, our data is spoken for with a language fashioned from these models. […] systems of classification algorithmically authenticate us online"* [10]. Meaning that data themselves might cause biases and that they need to be reinterpreted before they can build the basis for scientific discoveries: reasoning about them implies linguistic communication. Language in science requires a specific register and the sharing

of a common jargon, crucial to success in international research environments, as is awareness that the cultural background influences the meanings people read in words and actions.

The globalisation of science through the internet means not only that it is possible to communicate across continents: research data, technology - the digital revolution, the internet of things – impact on the representation of our own *selves* and the world. Floridi [11] underlines that "*we do not make science by mere accumulation of data*", so "*the real epistemological problem with big data is small patterns*. […] *the pressure* […] *is to be able to spot where the new patterns with real added-value lie in their immense databases*. […] *This is a problem of brainpower rather than computational power*". Since there is no dedicated academic path ready yet, "*philosophers might not only have something to learn, but also a couple of lessons to teach*" [11]. The *Harvard Business Review* describes the data scientist "*as a hybrid of data hacker, analyst, communicator, and trusted adviser*" [12]. To synthesize, different kinds of scientists are already *doing* data analysis and a multidisciplinary approach is advisable.

At international conferences like IRCDL researchers from different fields exchange views being displaced from their *safe havens*, obliged to explain anew issues that they are used at taking for granted, starting from nomenclature. False friends such as *library*, *collection*, *archive* evoke completely different images in the minds of a Library and Information Science (LIS) professional and of a computer scientist. The translation needed is not only to mediate between English and other languages, but also between the language of science and that of the humanities, to cross a chasm which has been perceived for a long time, at least since Snow defined it in his Reed Lecture [13] as a "*gulf of mutual incomprehension*". At a conference recently organised at the University "La Sapienza" upon the lecture's 60th anniversary, the theme of language surfaced in the challenges of defining the object of study and communicating science. In this respect, the name of Otlet [14] and the creation of the Universal Decimal Classification were mentioned: born of the urge to supply science with a further tool for its progress, his taxonomy was meant to assign a code number to disciplines, thereby helping highlight the relations between sciences, so that new science could be developed by the observation of the same object from different points of view. His idea was to overcome misunderstandings by means of the numeric notation.

For all that, the issue remains of what language is to be used within the scientific community, made increasingly global by the internet communication tools and the changing habits with respect to the publication process. Publication formats such as books and journal articles developed into preprints and raw data; their sharing happens via social media or, alternatively, in journals, institutional repositories or large open-access databases such as PubMed, ArXiv or the like. In any case, the language most widely used is English. PubMed homepage states that the database comprises over 30 million citations, but a language search returns over 25 million English records, whereas Chinese, Italian and Spanish score less than 400.000 records each, Russian over 600.000, French over 700.000. To date, a Scopus search by language yields 65,417,222 results in English, 1,413,945 in French, 751,903 in Spanish, 1,026,903 in Russian and 1,769,605 in Chinese. Though PubMed is US-based and open access, Scopus is a proprietary database produced by a Dutch corporation with a long, European history. Yet English rules in both cases.

Open and fair data implies a cultural translation; using languages other than English, also in contexts like *citizen science*, will promote *cultural justice* across nations, ethnicities, social strata, gender, age (the elderly, for example), foster the dissemination of the results of socio-relevant research, and hamper the potential imperialism of the English language. An example is MeSH, the medical thesaurus of the US National Library of Medicine, which, next to the medical jargon, enriches its Spanish version with the common language terms for illnesses: plurality of languages and registers.

2.2 Why English? English as a World Language

Pullum, co-author of the Cambridge English Grammar, while stating that the reasons do not matter anymore, and that we should simply accept English as a *lingua franca* and the advantages of having a common language at all, underlines that English did not gain this position because it was the language best suited for the purpose and "*it's rather important that people should realize, if they're English speakers, that this is their great piece of underserved good luck*" [15]. Wiener complained that Latin had failed as an international language, though it had once had the possibility of becoming "*an adequate international language far superior to the artificial ones as Esperanto*" [16]. English is here to stay, it works, therefore we are keeping the tool.

An effective strategy is a multilingual approach, as in the European Union "*mother tongue-plus-two*" principle enunciated in Barcelona, aiming at enabling "*citizens to be fluent in two languages in addition to their mother tongue*" and at raising "*awareness of the linguistic diversity of European society and turn it into an asset for intercultural dialogue and competitiveness*" [17]. This second goal is particularly meaningful in a multicultural and interdisciplinary research environment, moreover the idea is relying on the preceding UN choice of using more official languages - besides English also French, Spanish, Russian, Chinese, Arabic. Other international organisations, as IFLA, follow the example.

2.3 English as the Language for Science

Being an international language does not necessarily make English a language for science. This further development is definitely rooted in the history of science around World War II, the European diaspora of scientists first from the Nazi regime and later from the Soviet bloc. It was the time of Big science, the time when Wiener criticized the habit of certain classicists to freeze language in time and forbid its development into an effective tool for communicating. He pointed out that "*the Greek language of the time of Aristotle was ready to compromise with the technical jargon of a brilliant scholar, while even the English of his learned and reverend successors is not willing to compromise with the similar needs of modern speech*" [16], so English had a long way to go. Gordin seems to agree with Pullum, that "*There's nothing about English that makes it intrinsically better for science than any other language.*". He foresees that it "*could split into three languages: English, Chinese and another language, such as Spanish, Portuguese or Arabic.*" [18], yet, it is the common language now, and it "*is perfect for science: it's precise and straightforward*", says Bosch Grau [19]. This is the

result of a process: scientists adapted to using English and provoked a reverse process of adaptation of the language to science, so that nowadays "*English has acquired a vocabulary for concepts and processes*" [12].

Without the adaptive process there may be no English for international communication at all. Interestingly, it does not seem to be intrinsically more refined than other languages; for example in French, German or Russian, the term "science" encompasses "*scholarship in a broad sense, including the social sciences and often also the humanities*" [18] while its meaning in English is narrower. Yet, "*scientific activity is communicated in a language. I do not simply mean 'in words'; I mean in a particular, specific language, shared by a community of speakers*". Gordin distinguishes between *identity* – speaking of our inner feelings, in our native language, in different contexts - and *communication*, when we speak to be understood by a specific audience. "*If you are a native speaker of English, your language of identity equals your language of communication; your burden is reduced to the irreducible problem of saying what you mean*." [18]. Non-native speakers face a harder task.

Nature published an article following an incident at a US University, when a professor reproached some Chinese students for speaking their own language on Campus. Seven scholars, selected for their personal or professional experience with language barriers, offered their views on English as a *lingua franca* for science [19] and referred of episodes which witness to the advantages and disadvantages of English for scientific communication. Most agree that as a tool it is well developed: Chinese is a rich language, but "*it still lacks much of the vocabulary that's needed to describe physical science*", says Cheng, Physicist at Cornell University, for whom it would be difficult to give a presentation on her research topics in her native language. Elsewhere [20] she declares that it was precisely her English competence which allowed her to "*science*" her way out of her country and of a limiting social background. She remarks that a low English competence is often confused with a lack of skills or of clarity of mind, and witnesses that Chinese researchers who are not fluent in English do not feel like trying her path. The same happens for Spanish speakers, and, as Clarissa Rios Rojas underlines, that it is not just a matter of reading and writing literature in English, but of understanding the process and culture of science: this is why she founded a mentoring program, Ekpa'palek, "*that helps students from Latin America to navigate academia*" [19]. The same seems to be on the mind of Montserrat Bosch Grau, from Spain, presently Director of *in vitro* studies at Sensorion in Montpellier, France, who claims that to master the language means to advance both in research and on the job market. She worked hard to raise her English competence to a good level while studying 12 months in France when English was not in the syllabus. Vera Sheridan, Language and intercultural relations researcher at Dublin City University, is of Hungarian origin and understands what it is like to learn a foreign language. She compiled a list of resources for foreign researchers from all over the world and points out that language skills need to be taught, they cannot be implicitly absorbed [19].

2.4 The Contribution of Multicultural Research Perspectives

Relying totally on English for scientific communication could cause *open* and *fair* science to miss some opportunities in understanding the world diversity. The concept

of *open* was asserted through the BBB Declarations between 2002 and 2003, the concept of *fair* stands for *Findable, Accessible, Interoperable, Re-usable*, as synthesised in 2011 by FORCE11 [21]. If *open* and *fair* have become the pivotal principles of the Digital Libraries in our days, this is due to the redefinition of values brought about by the Post-Modernist thought and particularly due to the kind of critical perspective proposed by the Post-Colonial, Cultural and Gender Studies. A multicultural and interdisciplinary community, as the one represented at IRCDL, will seize on these values and integrate them in their *technical* choices.

Diversity. The perspective of the UN Millennium Goals regarding diversity is that problems must be faced globally, whereas non-English research is more difficult to spot and read out. When studying issues linked to local territories, the international community might miss out quite a few points. Irawan, hydrogeologist at the Bandung Institute of Technology, following the example of the first-ever open world repository, arXiv, founded *INA-Rxiv* [22] to increase visibility of papers in Indonesian – the language of one of the countries with the highest biodiversity in the world - that describe issues of geology, biodiversity, geography linked to that territory.

Tatsuya Amano [19], Zoologist at the University of Queensland, Brisbane, after realising that 36% of over 75,000 biodiversity conservation papers published in 2014 in PLoS Biology were in languages other than English, therefore less accessible, started a research for non-English literature in the field, struggling to have it translated into English. In other words, "*Just because something's not in English doesn't mean you should ignore it*" [23], says Bond, a senior conservation scientist at the Royal Society for the Protection of Birds in Sandy, UK.

Field research would benefit from a multilingual approach. Shena Dharwadkar, from the Centre for Wildlife Studies in Bengaluru, India, is a herpetologist and remarks that scientists from Western countries prefer to hire English-fluent guides to carry out field research in her area, yet, she underlines, "*locals understand the problem better*" [19], no matter whether they speak English or not. Moreover, as she points out, science should reach out to residents and be beneficial to everyone. Wildlife researcher Owen Bidder [23] explains that, being proficient in German, he can get precious information from German hunters during his field research.

Managing Interaction and Career. Much research is led globally, in large laboratories, involving international, multicultural and multilingual research groups, where every member has different tasks. To work well together, they must feel comfortable: Chen [19] observed that normally European researchers speak in their native language, whereas Chinese and South-Koreans just do not feel comfortable at doing the same. Conversely, Clarissa Rios Rojas seems to feel that being from abroad has some advantages, as she, being a native Spanish speaker, is able to relate easily with other Spanish, Portuguese and Italians. Having people build a community definitely facilitates working together and exchanging views, but it requires either a good level of English or a diffused multilingualism in the group and certainly an open attitude.

For non-English speakers, tasks such as writing papers and applying for fellowships are harder. Amano pointed out that "*The dominance of English has created considerable bias in the scientific record*" [19]. A difference in perspectives is enriching to the scientific discourse, and a low competence in English – though it might still allow to be

speakers at conferences - makes it harder to network, which is the main point of attending conferences, after all.

Moreover, submitting articles in English opens the path to core journals, the ones indexed in large databases, Scopus and WoS, and consequently to the highest evaluation, impact factor and worldwide visibility. The language issue is but one ring in a whole chain of research assessment and publication policies, regardless of the outlet: journals, proprietary databases and open access repositories, so an issue which might look *simply* language-related refers instead to the stabilization or to the rethinking of a whole system of values.

Cultural Context and Exchange. The physical and linguistic position of a digital collection is not irrelevant, suggests Castellucci [24]: ArxiV "*tastes*" of Los Alamos and New Mexico, its history can be read as part of the cultural context it developed into. Moving from the assumption that there is a *materiality* inherent in digital resources, one of the main concerns of policies for the development of Digital libraries nowadays regards the *decolonization* of electronic archives - in other words, a conscious effort in the selection of languages, collections, corpora as well as literary and aesthetic canons other than those mediated through English language [25]. "*English speakers have become the gatekeepers of science. By keeping those gates closed, we're missing out on a lot of perspectives and a lot of good research*" [19], says Vera Sheridan. Research needs to be led from different viewpoints, "*If you're a non-native speaker, you can bring a diversity of opinion and approach to the international community*" [19], says Tatsuya Amano.

Even the father of Cybernetics, Norbert Wiener, is extremely aware of the relevance of the cultural context, and often refers to the one he works within, the USA in the Fifties. A look at some quotes clarifies what that time and place "*tasted*" of: "*The education of the average American child of the upper middle class is such as to guard him solicitously against the awareness of death and doom.*"; "*It is possible to believe in progress as a fact without believing in progress as an ethical principle; but in the catechism of many Americans, the one goes with the other*"; "*To the average American, progress means the winning of the West.*" [16]. There would be much to say on these statements from an historical and social point of view, and it would be beyond our scope to go deeper, but it might be surmised at least that, at the time, the education of the average German or Japanese child was based on totally different experience. Moreover in Wiener we find the same *taste* of American society in the concept of *information* itself - "*The fate of information in the typically American world is to become something which can be bought or sold.*" – and in the idea of research as Big science, when he says "*The skill by which the French and English do great amounts of work with apparatus which an American high-school teacher would scorn as a casual stick-and-string job, is not to be found among any but a vanishingly small minority of our young men.*" [16]. Wiener tells us of a society where research is supported by large funds because it is a product for the market. In fact, he laments the lack of artistic creativity and proper training in the practice of science, the subservient attitude of research to the market, and feels that big laboratories are over-structured for a future where he does not "*foresee that the next generation will be able to furnish the colossal ideas on which colossal projects naturally rest.*" [16]. Moreover he reflects the "*taste*"

of a time when the scientist was male (*him*; *young men*), and *American* was synonymous with *from the USA*, when the perception in terms of *gender* and *post-colonial* had not settled yet – though even recently Isabel Allende was very surprised when an immigration clerk in the USA insisted that she should qualify herself as *coloured* and not *American* [26].

In a world of Big data we need all the different perspectives that we can use. Wiener defines himself as a "*scientific artist*" [16], because science is the language he chose to have his say, and he is most aware of the role of art and the humanities in the progress of human knowledge. We saw above that tasks and skills necessary for data analysis require creativity, since this is a completely new field we are still striving to understand. In a perspective of open, fair science the predominance of English should become less crucial, in order to prevent missing valuable publications and points of view only because the people expressing them do not master the language.

3 The Role of the Librarian and of the Humanist

Much boost to research is given by the possibility to access to raw data, pre-prints and publications in an open and fair way from wherever, and to be able to share views with researchers from different cultures. The contribution of specialists from different disciplinary backgrounds to perform specific tasks is beneficial to the work of a research team. Among them there is information curation and organization – something library and information specialists have been taking care of for a long time.

A digital library is defined by IFLA as "*an online collection of digital objects, of assured quality, […] created or collected and managed according to […] principles […] and made accessible in a coherent and sustainable manner[*which*] forms an integral part of the services of a library*" [27]. The traditional mission of the librarian transposed into the digital environment implies collection curation, organization and access to materials for the specific communities served. The role of a librarian within an academic organization or a research team is exemplified by the experience of Luisella Goldschmidt-Clermont, who was for ten years Senior Scientific Information Officer at the European Organization for Nuclear Research (CERN) within the community of high-energy physicists [28]. She remarked that scientists, absorbed by their research duties, struggled to keep up with literature updates and with external communication of their results and suggested that some members on the research team should be specifically in charge of that. Differences are perceivable between scientist and humanist: Goldschmidt-Clermont observes that physicists are not keen on allocating more funding and staff to information management and communication. This might be a mark of the difference perceived between STEM and humanities scholars, especially in the digital world, where mathematicians and the like are at home, and where the humanists arrived later, as guests [24].

The competences of librarians and information specialists mediate between hard sciences and the humanities; LIS curricula combine the world of printed items with Information Science and Technology. It could not be otherwise, since the organization of documents passes through their translation into records, granular [29] representations of real materials, whatever their format. It was not by coincidence that CILIP, the

UK professional association, sprung in 2002 from the merger of the Library Association and the Institute of Information Scientists. Since the librarians' mission is to "*organize and present content in a way that allows an autonomous user to find the information s/he needs*" and "*to provide the best possible access for library users to information and ideas in any media or format."* and that "*This includes support for the principles of open access, open source, and open licenses*." [6], they would naturally be interested in any format of information. As a result, if the information goes digital, they need to become more proficient in IT.

After the digital revolution, being able to organize information in the best way has become crucial to the survival of science itself. Big data challenge libraries "*to adopt new service models to assist with the transformation of data into information*" […]. "*Today data science is seen as the blending of competencies in computer programming, software engineering and statistics, combined with a particular domain expertise*" [30]. The data librarian is supposed to be able to work closely with researchers, yet the tasks and skills of this professional figure are still being defined. Library associations and academia are behind in preparing standards for the job – the exception is CILIP, stating that data librarians are "*engaged in managing research data, using that data as a resource and supporting researchers in these activities*, and to that purpose they are likely *to be involved in developing or implementing an organization's Web data management plan, storing and managing data and determining retention and disposal periods*" [31]. Among qualifications, apart from a LIS degree, "*understanding of current technologies and standards such as institutional repositories, encoding standards (e.g. XML) and metadata*" are recommended.

Apparently, as in the case of the data scientists, specialists are already out there doing the job, but specific education is not ready. Moreover, tasks are all but defined. Recently, a research on "*What is a data librarian*" [32] analyzed job ads in the USA and found that the most required skill is research assistance, followed by a critical thinking/problem solving attitude. Apparently, the main task envisioned for the data librarian is instructional support, including copyright, intellectual property, licensing of data, embargoes, ethics and reuse, data literacy and privacy.

Data librarians are used at speaking "*multiple disciplinary languages*" [30], and therefore can play a role as connectors in a multicultural research environment; moreover, they are familiar with classification systems, which serve as multidisciplinary links, as Otlet [14] pointed out. Nevertheless, such a role in an international environment demands language competence. The right attitude, therefore, might be to raise awareness on the fact that English is a tool for communicating, while bearing in mind the value of vernacular, local literature. Librarians, who manage collections by standards, languages of their own, could help communicate internationally the products of research in local languages and the cultural diversity they represent. Moreover, learning a language is not about finding different labels for the same boxes, as in different contexts it is the boxes and their contents that change. Learning technical English and discussing the terminology for library services in another language "*enhances the librarians' perception about their role*" and helps them remain "*inventive and creative to match the transformations in the nature of information and society.*" [33], states Fontanin, who has been teaching English to librarians for the past twenty years. Foreigners coming to the libraries might range from refugees to visiting

researcher: the variety in meanings, contexts, social scenarios is as large as science, as diverse as society.

4 Attitude for Solutions

Access to multilingual, multi-register and multidisciplinary information fosters the development of a free science, but it is not always available. As Hibert [34] recently reasoned, in countries where funding is low and English proficiency not that high, Shadow Libraries [35] thrive, a bottom-up solution to the access to knowledge. Piracy according to some, resistance to others, they witness a critical issue in an apparently universal digital era: across embargoes, Internet blocks and expensive subscriptions, looking at the world from a non-Western perspective shows a not-so-fair landscape.

Probably the best approach is to start small, from practical solutions to single problems. Some examples are proposed by the interviewees on Nature [19]. First, English language should be taught at University level, specifically for those who are undergoing a research path. In addition, mentoring should become a habit: Vera Sheridan claims that writing experts are not enough to turn a PhD thesis into an article, as the process demands specific competence in the discipline. Extending the *embedded* approach – librarians and teachers working side by side to reach course goals - to larger research groups could prove beneficial.

International organizations should provide specific allocations to language support in the projects they finance. Shena Dharwadkar proposes that hiring locals to assist researchers in field research should be made a criterion to privilege project funding, and the same could be said for the mediation and dissemination of research in local languages with the help of libraries, data librarians and scientists.

5 Conclusion

All in all, though English is a very useful tool to exchange views in a global scientific community, a multilingual approach would not only facilitate conversations, but also envision different points of view. Learning a language is about experiencing a different worldview, it is a transformative experience, as is research.

In a research group diverse points of view enrich the final product: local perspectives, individual views, multidisciplinary approaches may open up new visions, therefore the role of women, of minority languages and cultures, of digital humanities contributes to creativity. Wiener was the father of cybernetics and defined himself as a scientific artist. Provided the art offered is aligned with the spirit of global science and with people's needs, provided it is striving to convey a message, it does not matter what language or alphabet is used to express it, be it text, visual, or else. An art installation by Dormino [36] used art to make a point on Web information: between 2015 and 2017, four chairs were placed in various locations. Three controversial figures - Edward Snowden, Julian Assange and Chelsea Manning - were standing on them, a fourth was left empty for whoever had got "*anything to say*". It is another language, but it is effective.

We could close in the words of a scientist and a humanist. Albert Einstein stated: "*Man tries to make for himself in the fashion that suits him best a simplified and intelligible picture of the world; he then tries to some extent to substitute this cosmos of his for the world of experience, and thus to overcome it. This is what the painter, the poet, the speculative philosopher, and the natural scientists do, each in his own fashion. Each makes this cosmos and its construction the pivot of his emotional life, in order to find in this way peace and security which he cannot find in the narrow whirlpool of personal experience*". Looking for answers requires a multi-faceted approach and a multitude of languages. This richness would be *lost in translation* - in the words of Roman Jakobson, founder of modern linguistics - if it were channeled through a single language code.

References

1. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003). https://openaccess.mpg.de/Berlin-Declaration
2. Fontanin, M., Castellucci, P.: Water to the thirsty reflections on the ethical mission of libraries and open access. In: Manghi, P., Candela, L., Silvello, G. (eds.) IRCDL 2019. CCIS, vol. 988, pp. 61–71. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11226-4_5
3. Ross, A.: Real Love: In Pursuit of Cultural Justice. Routledge, London (1998)
4. United Nations: Universal Declaration of Human Rights (1948)
5. Shah, H.: How to involve citizen scientists to develop AI for Knowledge Exchange. Presented at the OAI11 – CERN-UNIGE Workshop on Innovations in Scholarly Communication, 19–21 June 2019
6. IFLA FAIFE <Freedom of Access to Information and Freedom of Expression>: IFLA Code of Ethics for Librarians and other Information Workers (2012)
7. The Joy of Stats. https://www.gapminder.org/videos/the-joy-of-stats/
8. Chan, L.: Contextualizing Openness: Situating Open Science. University of Ottawa Press, Ottawa (2019)
9. Budapest Open Access Initiative—Read the Budapest Open Access Initiative. https://www.budapestopenaccessinitiative.org/read
10. Cheney-Lippold, J.: We Are Data: Algorithms and the Making of Our Digital Selves. New York University Press, New York (2017)
11. Floridi, L.: The 4th Revolution: How the Infosphere is Reshaping Human Reality. Oxford University Press, Oxford (2014)
12. Davenport, T.H., Patil, D.J.: Data Scientist: The Sexiest Job of the 21st Century, HBR, October 2012. https://tinyurl.com/jb4lgum
13. Snow, C.P., Collini, S.: The Two Cultures. Cambridge University Press, Cambridge (2014)
14. Otlet, P.: Traité de documentation. Le livre sur le livre, théorie et pratique, par Paul Otlet, Bruxelles, Palais mondial (1934)
15. Pullum, G.: UQx Write101x English Grammar and Style - Interview with Geoffrey Pullum. https://tinyurl.com/sros83w
16. Wiener, N.: The Human use of Human Beings: Cybernetics and Society. Eyre and Spottiswoode, London (1950)

17. Commission of the European Communities: Accompanying document to the Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Multilingualism: an asset for Europe and a shared commitment (2008). https://tinyurl.com/skd4gbq
18. Gordin, M.D.: Scientific Babel: How Science Was Done Before and After Global English. The University of Chicago Press, Chicago (2015)
19. Woolston, C., Osório, J.: When English is not your mother tongue. Nature **570**, 265 (2019). https://doi.org/10.1038/d41586-019-01797-0
20. Cheng, Y.: Let science be science again (2017). https://tinyurl.com/ug95gy9
21. FORCE11. https://www.force11.org/
22. Irawan, D.E.: Against all odds: we only see numbers. In: Indico (2019)
23. Walker, C.: Learn the local lingo to get ahead. Nature **534**, 425–427 (2016). https://doi.org/10.1038/nj7607-425a
24. Castellucci, P.: Carte del nuovo mondo: banche dati e open access. Il mulino, Bologna (2017)
25. Di Loreto, S., Morello, S.: Decolonizing the digital archive. In: 25th AISNA Biennial Conference - Gate(d)Ways. Enclosures, Breaches and Mobilities Across U.S. Boundaries and Beyond., Ragusa (Italy) (2019)
26. Allende, I.: Mi país inventado. Areté, Barcelona (2003)
27. IFLA – IFLA/UNESCO Manifesto for Digital Libraries. https://tinyurl.com/rmwbw5l
28. Goldschmidt-Clermont, L.: Communication patterns in high-energy physics. High Energy Physics Libraries Webzine, no. 6 (2002)
29. Zani, M.: Granularità: un percorso di analisi. DigItalia **2**, 60–128 (2006)
30. Klapwijk, W.: IFLA. Big Data Special Interest Group: A concept framework for data science in libraries (2018). https://tinyurl.com/vbg73c6
31. Data Librarians - CILIP: the library and information association. https://tinyurl.com/sue8nql
32. Khan, H.R., Du, Y.: What is a Data Librarian? Content Analysis of Job Advertisements for Data Librarians in the United States Academic Libraries (2017)
33. Fontanin, M.: The flexible Librarian: English @t the Circulation Desk. AIB, Roma (2017)
34. Hibert, M.: Ecologies of smart unstructuring: silicon regimes, alternatives in commons, and unparallel librarian. Presented at the IFLA Satellite Librarians and Information Professionals as (Pro)motors of Change, Zagreb, Croatia, 20 August 2019
35. Karaganis, J. (ed.): Shadow Libraries: Access to Knowledge in Global Higher Education. The MIT Press; International Development Research Centre, Cambridge; Ottawa (2018)
36. Dormino, D.D.: Anything to say? A monument to courage 2015–2017. https://tinyurl.com/tml22td