



ASSUREDSCIENCE

Discover how and why Assured Science came to be

7: Clues used in designing the new system for research

The task was to create a satellite system that would work in the most democratic, efficient, inclusive, versatile, and vibrant manner possible for generations to come.

I have never used EndNote or any other citation manager. In part due to the boring chore of establishing the base library after being in the field for more than a decade. And in part because hard copies of all the publications I reference regularly were strewn or piled up on my desk or bench in a seemingly chaotic manner that made perfect sense to me. But it was mostly because I remembered the first and the last authors of all the publications that contained data relevant to my research. On some of the older publications with few authors, I remembered all of them. Therefore, it was easy enough for me to type in the new citations or cut and paste old ones to create citation lists for publications or grant applications. Of course, I had to deal with the different formats for different journals but that was an easier chore.

Anyway, I even used to ask questions in seminars citing data using author names and the year of the publication. I began to understand this penchant

1. More time for research: Fund people not projects
John P. A. Ioannidis
Nature 477, 529–531 (29 September 2011)
2. Dr. No Money: The Broken Science Funding System
Scientists spend too much time raising cash instead of doing experiments
The Editors, *Scientific American* 304, 10 (2011)

Figure 1

of mine when I started to think about the problems in the traditional system for research. Now I think that most scientists must have this penchant, whether they are conscious of it or not and whether they show it or not, because *the system cultivates it*. Scientists essentially own their data, even after publication (hence the necessity for citation). They are even encouraged to become possessive of their data because data that are described as

'groundbreaking,' 'cutting edge,' or 'innovative' bring them rewards in terms of grants, jobs, opportunities, and reputation. Scientists producing such premium data are considered to be wellsprings and the presumption is that more such data would flow from them. It is because of this perspective that some people suggest to just fund scientists in order to simplify funding issues (Figure 1). Keep in mind that the scientists are the primary focus of

the traditional system and any desired effect on the mechanism or data is mediated indirectly through actions focused on the scientists.

Problems related to funding, mechanism, and data have become so acute that scientists, science administrators, and science journalists have written



numerous articles over the last couple of years discussing ways and means to fix the traditional system for research. Some of them focus on many aspects of the system and have suggested many 'fixes' for improvement. Two such articles are shown in Figure 2. They have served the dual purpose of highlighting the system wide problems and call for action.

Suggestions in these and other similar articles are all sound and seem plausible to effect the targeted changes, but I do not think they are very probable to do so, at least not to any significant degree. The reason is that they are based on unrealistic assumptions or expectations and ignore, even reinforce, one or more of the myriad of problems facing the traditional research system. For example, one proposal is reduction in the size of grants and placing limits on the number grants awarded to a single scientist so that more eligible applications can be funded. This proposal ignores the potential for a corresponding decrease in research quality, increase in uncompleted projects, budget inflation, and/or penalizing productive scientists.

Another proposal is for scientists and institutions to exercise self-control and self-policing. Obviously, the call goes to those that are successful or very successful, but why would they heed the call? Is it even reasonable to ask the successful to be less successful? In my view, most of the proposed solutions depend on the sacrifice and idealistic behavior of those that flourish within the system, or expect changes in the current research culture without incentives or penalties. Remember that in the traditional system, all changes have to be effected through people, and people will be people. They can be expected to maximize their success and benefits as much as possible, within the bounds of normal and ethical behavior.

Many other articles focus specifically on funding problems. A sample of them is shown in Figure 3. As the federal government provides the most funds for



science in the United States, more than a three-fold increase is necessary to fund all viable science projects (i.e., at least about 70% of scientists). I do not think this is going to happen in the foreseeable future given the existing political and socio-economic climate. Even in the remote possibility that were to happen, it would only lead to more inefficiency and waste if the other mechanisms and data problems are not mitigated. Some proposed

funding initiatives target specific segments of scientists, such as young people (see Figure 3.3). In my opinion these are based on anecdotal or even erroneous assumptions.

For instance, I am sure there are younger scientists discovering groundbreaking data, but I bet there are also older people doing it. I also disagree with the assumption that younger scientists would be in general more successful and productive than older scientists. They might take higher risks and approach a problem with more irreverence but these attributes are generally more useful when a field is in the discovery phase. In the more mature or complex fields (e.g., biology) experience and cumulative knowledge is also extremely useful for efficient progress. For example, with more than 25 years of bench experience and first hand knowledge of primary data on my subject, I am confident of out-producing and out-thinking most young people in the initial stages of their independent scientific career researching my subject, the *Notch* gene.

Of the many mechanism problems in the traditional system, only wastage of time as a consequence of writing so many grant applications has received any serious attention (for example, see Figure 1). In my opinion, the proposed solution, to identify good scientists and fund them (that is basically what organizations such as Howard Hughes Medical Institute do) is not as easy as it seems. How can one identify good scientists who will continue to produce good data for more than 30 years? The past is merely an indicator, not a guarantor of future success, and what is considered as success itself might change in the future. Furthermore, such an approach will not only make the system undemocratic but also aggravate all the data problems.

Other mechanism problems that contribute to inefficiency, gaming, and exclusion in the traditional system have not been considered or have been merely acknowledged without any actionable proposals to address them.

Data problems, on the other hand, have received considerably more attention and potential solutions have been seriously discussed. A sample of articles addressing some of the data problems is shown in Figure 4. When

1. Policy: NIH plans to enhance reproducibility.
Francis S. Collins & Lawrence A. Tabak
[Nature](#). 2014 Jan 30;505(7485):612-3
2. Drug development: Raise standards for preclinical cancer research
C. Glenn Begley & Lee M. Ellis
[Nature](#) 483, 531–533 (29 March 2012)
3. Power failure: why small sample size undermines the reliability of neuroscience
Katherine S. Button, John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson and Marcus R. Munafó
NATURE REVIEWS | NEUROSCIENCE VOLUME 14 | MAY 2013 | 365

Figure 4

thinking about data, it is important to keep in mind that there are two components to it: precision and accuracy. Precision refers to reproducibility; from a statistical perspective it means low error variance. In other words, minimizing the influence of all factors except those under study so that the same result is reproduced whenever the experiment is repeated. All the proposals in articles such as those shown in Figure 4 basically address precision. There are now even commercial companies offering services to verify or validate data. The proposed means for

improving data in articles such as those shown in Figure 4 are good and should be implemented to the maximum extent possible. However, I do not know how that can be accomplished for all the data in the system given the idiosyncratic way most laboratories function, or who is going to bear the cost of validation by external agencies. I can envision investment by someone interested in some portion of data with the potential for commercial value, but what about the majority of data that are also important for progress?

In any case, I am more concerned about the second component of data, accuracy, which I consider to be the bigger problem. Accuracy refers to how well data explain their respective phenomenon or observation. This is a very difficult issue to either assess or rectify in the traditional system, as the majority of data are idiosyncratic, disparate, and biased in favor of those that fit popular models or can be otherwise explained. The notion that depositing unpublished orphan or negative data in accessible websites would resolve accuracy problems is unrealistic to say the least.

For one thing, unpublished orphan data would be even more disparate (as related data would have been published elsewhere) and extremely variable in quality (a primary reason for not being included in a publication). Furthermore, as the data in the system are so idiosyncratic only the person or the laboratory that generated them would really be able to use them effectively. In addition, when published data that have gone through the review process have a low chance of being directly integrated into someone else's research, what are the chances for unpublished, orphan data deposited in a website? Why would someone give up the data if they are still writing grant applications? Why would someone spend the enormous amount of time and effort processing data that is necessary for meaningful searching

and easy retrieval from a website? To my mind, it is not like depositing gene sequences. One has to provide the context, methods, materials, and interpretation for all data deposited, which is like preparing data for publication. And what do scientists get in return for deposition? If data deposition is forced, what do others get from merely 'dumped' data?

The second thing is, I do not know what to make of the call for depositing negative data. It is possible, and is already done, for data with high human or economic value (e.g., clinical trial data). I am concerned about other kinds of basic and applied data. From the bench perspective, unless one is working on a simple deductive project, the majority of data obtained (maybe even 90% in some cases) are basically negative data.

In my view, accuracy can be seriously addressed only when data in a field are continuously integrated in an unbiased manner, which is simply not possible in the traditional system wherein any effect on the data has to be mediated through actions on a biased and small sample of scientists selected for funding using a mechanism with serious problems.

Recently, considerable attention has been paid to scientific fraud. Three articles in as many weeks in a prominent newspaper, including an Editorial,



should give you an idea of the level of attention it has received (Figure 5). Both journals and funding agencies have developed elaborate tools and procedures to catch fraudulent scientists and penalize them. Even commercial watchdog companies are cropping up to scrutinize data suspected to be fraudulent and develop evidentiary material for penalizing the scientists involved (Figure 5.3). Scientists are no different from any

other professionals and should be penalized if found cheating. But I think the zeal and cry about its impact on scientific research is out of proportion to the scope of the problem.

Estimates of fraudulent data (excluding inadvertent errors or sloppy data) are in the 1-2% range (Figure 5.3), which would be (from a statistical perspective) considered insignificant in most situations. Yes, this 1-2% data consumes money, but I think eliminating fraud completely, if it were at all possible, would be even more expensive. I do not know of any human profession that is free of fraud and I do not see the reason why scientific research should be any different. When it comes to ethical behavior, I do not think scientists are any different from others — most scatter around the norm, some rise above it, and some fall below it. Given the idiosyncratic and

disparate nature of data, I think any additional increase in policing measures would hurt research within the traditional system. The reason is that the vast majority of scientists (I would say more than 95% or them) conduct research in good faith, genuinely believing that they are addressing an important problem, and do their best given the resources available to them.

The notion of a well-designed experiment perfectly executed to answer a question is wishful thinking in most situations. The reality is an error-prone and messy process that requires an attitude composed of, in equal parts, the confidence to pursue an idea and the humility to learn from mistakes. This attitude would be stifled in a 'police state of research.' But let us imagine that policing and watchdogging increase. They might net more of the egregious data manipulators but barely impact data on the whole. That is because policing would affect only a very small fraction of data (the 1-2% that are fraudulent) and it would impact mostly the precision component of data, not accuracy.

Showing that data do not reproduce well (i.e., lacks precision) is trivial compared to showing that they do not fit a phenomenon or observation (i.e., lacks accuracy). Keep in mind that the scientist who generated the data is the one with the most detailed knowledge of the subject matter, which is often complex. I am not arguing for condoning fraudsters and cheaters. I simply think there is a better way to address the problem more effectively (which is incorporated into the new system for scientific research).

After considering all the major proposals for fixing the traditional system for research, I realized that the funding problem is not going to ease in any foreseeable future, which means that the 50-60% of additional projects that could be funded would not be funded. As a consequence, all the scientist talents and scientific resources associated with these projects would be essentially wasted. It was also clear to me that most of the mechanism and data problems I see in the traditional system would not be resolved by the proposals put forward by other scientists, administrators, and funding agencies. I spent several months trying to develop more fixes to the traditional system that would mitigate the unaddressed problems. I failed.

I was on the verge of giving up when out of desperation I started wondering what I would do if I were tasked to design a new system for research that eliminated or minimized all the problems I see in the traditional system. It was then I realized that the basic problem with the traditional system is that it is focused on scientists instead of on data, despite the fact that in general data outlast scientists. I started thinking about designing a new system for research with data as the focus and doing what is best for data in a field in the long run. The condition I put forth is that the funding, mechanism, and

data problems observed in the traditional system must be eliminated, mitigated, or avoided using tools that work with the best and most positive aspects of human nature and behavior.

Here is the list of the criteria or guidelines that I used for designing the new system for research.

- There should be equal opportunity for funding with the ultimate goal of funding all competent scientists interested in scientific research.
- Scientists should have the freedom to pursue any scientifically viable project in which they are interested (basic, applied, amusing, etc.).
- The mechanism should be as efficient as possible with minimal waste of time, talent, data, and resources.
- The data should be very robust, integrated, and organized in a manner that enables maximal and efficient use.
- It should work within today's funding constraints and within current rules and regulations governing research.
- It should provide unlimited potential for growth and support for both scientists and science.
- It should use natural motivations of people to provide stable solutions to mechanism and data problems.
- It should support even adventurous and risky projects.
- It should be flexible and adapt to changing conditions or advancing technology.
- It should be attractive to all: scientists, institutions, funding agencies, politicians, and the public.
- It should incorporate or respect lessons learned from the history of science:
 - Groundbreaking data often emerge in the fringes of a field;
 - Rapid progress occurs when both scientist number and project diversity are high in a field;
 - Delay in progress is due to non-intersection of data from different scientists;
 - True significance of data is often obscure for long periods of time or changes over long periods of time;
 - Peers poorly predict who among them is the next groundbreaking scientist (extreme example: Darwin, who was also working on inheritance, apparently failed to appreciate Mendel's data).

I have managed to develop the concept for a new system for scientific research that fulfills the above criteria and guidelines. It is designed to function as a satellite system to the traditional system, focused on

supporting and empowering bench scientists (i.e., scientists who generate primary data) and small laboratories. I have striven to see that the concept yields a system that would work in the most democratic, efficient, inclusive, versatile, and vibrant manner possible. I believe that this system would also restore and keep the romance of scientific research in the generations to come.

Several people who have listened to my concept strongly suggested that I go the 'for-profit' route, as it would be easier to raise money. I refused and took the non-profit route because I think that scientists, science, and stakeholders committed to the best interest of scientific research should always be in control of the organization. I believe that they will always do what is best for science and scientists, not profit-oriented stockholders.

The concept includes components from diverse fields. While I am confident that each of these components individually and all of them together would work as expected, the involvement of more scientists from diverse fields and specialists in different components of the concept would vastly improve the capacity and potential of the system. I invite those of you who are passionate about scientific research, and have the time, talent, or money, to become involved. Although I am launching this effort with my own money, I seek no personal benefit. The patent rights and all copyrights are assigned to the organization. Whatever I do is for the benefit of the organization.

I think an initial team of 10-12 co-founders (including me) and a similar number of advisors would be able to translate the concept into an effective new system for research. Getting this system to the starting line and recruiting the pioneering group of researchers to begin conducting research in it is my only goal for the remainder of my life.

I am looking for people who are action-oriented and who passionately believe in the cause. You do not have to agree with all aspects of the concept or with me as long as you are willing and able to provide a better alternative for us to pursue together. My only dogma is doing what is best for scientists and science under the current circumstances and in the long run. I guarantee to each one of you who is interested in becoming a co-founder, advisor, or backer that I will consider your response with utmost consideration.

Please understand my request for specifics regarding how you plan to contribute to the organization, your CV, and your assurance of confidentiality before divulging more details of the new system. They are needed to protect the concept, keep the focus firmly on action, and to know from your background what expertise you bring to the effort.

If you are a potential candidate for being a co-founder, an advisor, or a backer (financial or otherwise), I will provide further details in person at a meeting set up in a mutually convenient place anywhere in the United States. If for some reason you are unable to travel to a mutually convenient place, I will visit you at your place if you seem to be a good fit. We will include your Essays in the Founding page and start our partnership with a handshake or a hug. We would need it to overcome the only hurdle I see: getting people to step 'outside of the box' and objectively assess what we are capable of accomplishing for the health and future of scientific research.

Consider the motto of the organization Assured Science Exchange, Inc.: *'More scientists and better science with less money.'* Everyone benefits: at least 50-60% more scientists can get funds to conduct research; institutions can collect about three times as much in indirect funds for developing infrastructure to support research; funding agencies can fund at least three times as many projects with the same money; politicians can avoid the pressure from constituents to appropriate more money for research; commercial companies can work with very robust and better data for developing consumer products; and the public can enjoy the maximum possible benefits from scientific research, including creation a new array of professional careers and avenues for economic activity.

The time is ripe for capitalizing on the most advanced and time-tested tools and devices available today to create not just a much needed system but also an irresistible one for charting a better course for the future of scientific research.

Cedric