TWIST BIOSCIENCE

Protecting DNA Data Storage:
Biosecurity for an Abiological Application

JAMES DIGGANS

Synthetic nucleic acids are dual-use materials

"BRIDGING THE DIGITAL TO PHYSICAL"

2002: Wimmer et al create polio *from scratch*

Chemical Synthesis of Poliovirus cDNA: Generation of Infectious Virus in the Absence of Natural Template

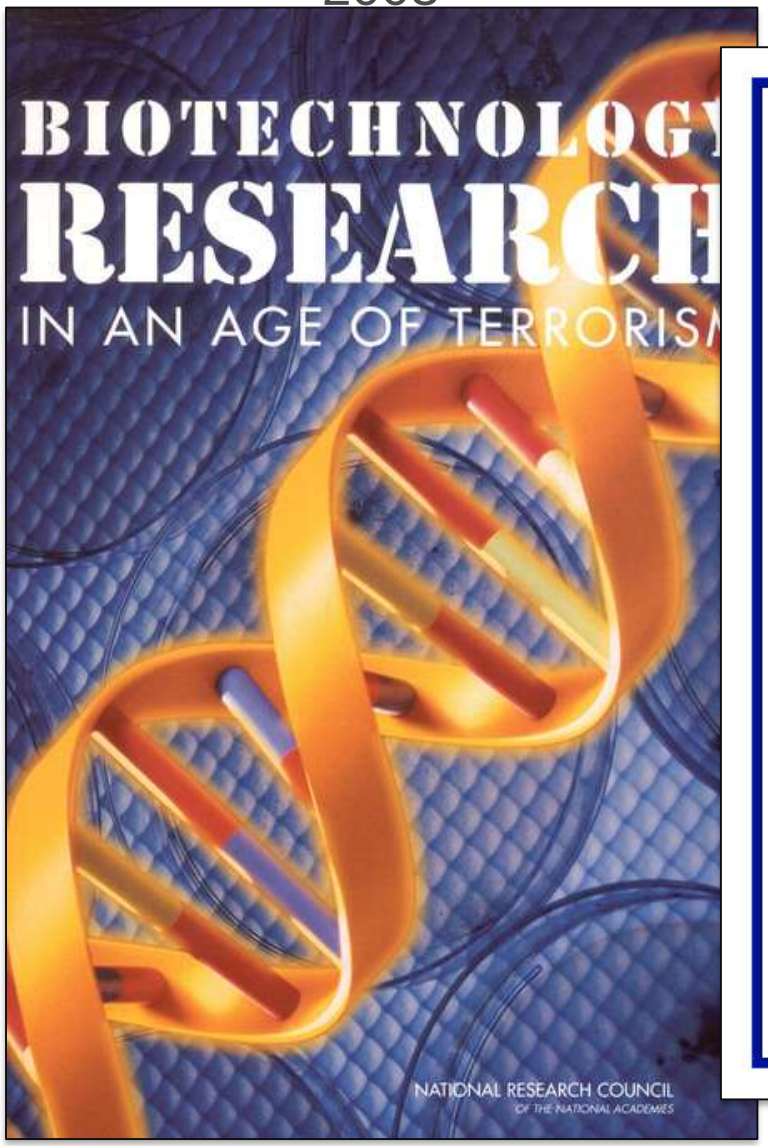
JERONIMO CELLO, ANIKO V. PAUL, AND ECKARD WIMMER Authors Info & Affiliations

SCIENCE · 11 Jul 2002 · Vol 297, Issue 5583 · pp. 1016-1018 · DOI: 10.1126/science.1072266

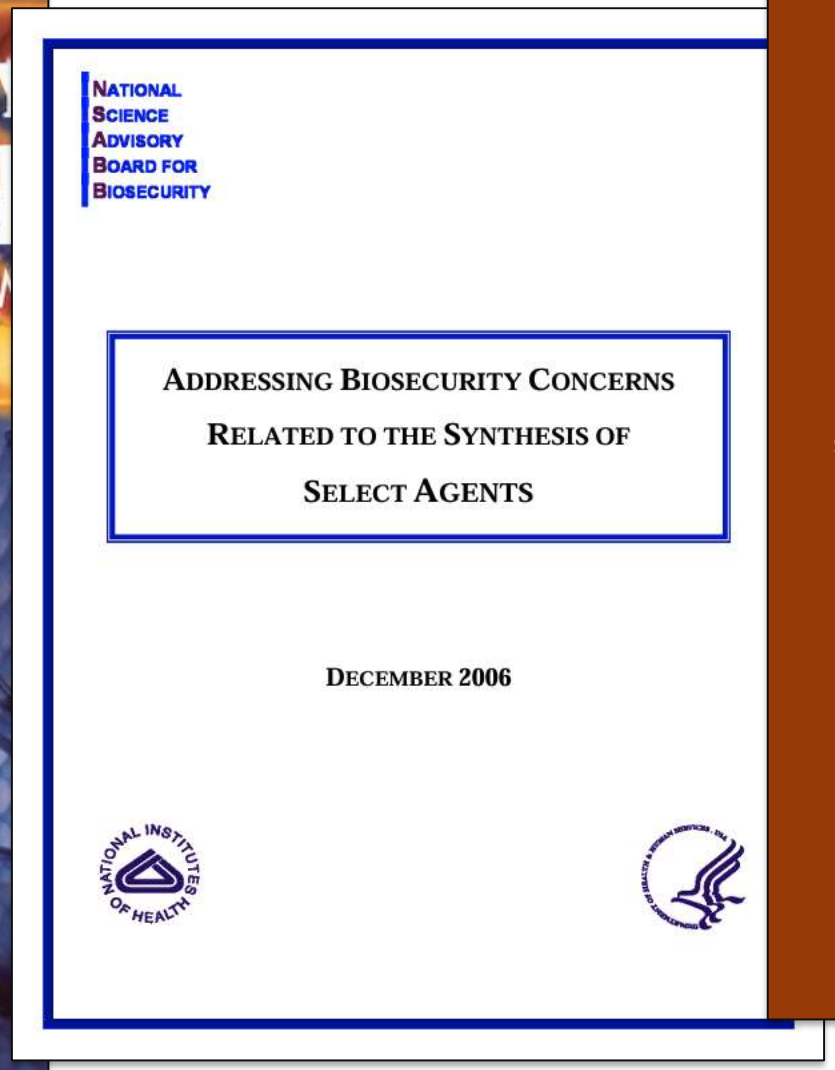
- Took 14 months to assemble ~7,500 bp
- Purchased oligos commercially and assembly services
- *During* the ongoing polio eradication campaign!
  - "The potential for virus synthesis is an additional important factor for consideration in designing the closing strategies of the poliovirus eradication campaign."

2000s – Rising Concern re: Dual-Use in Biotech

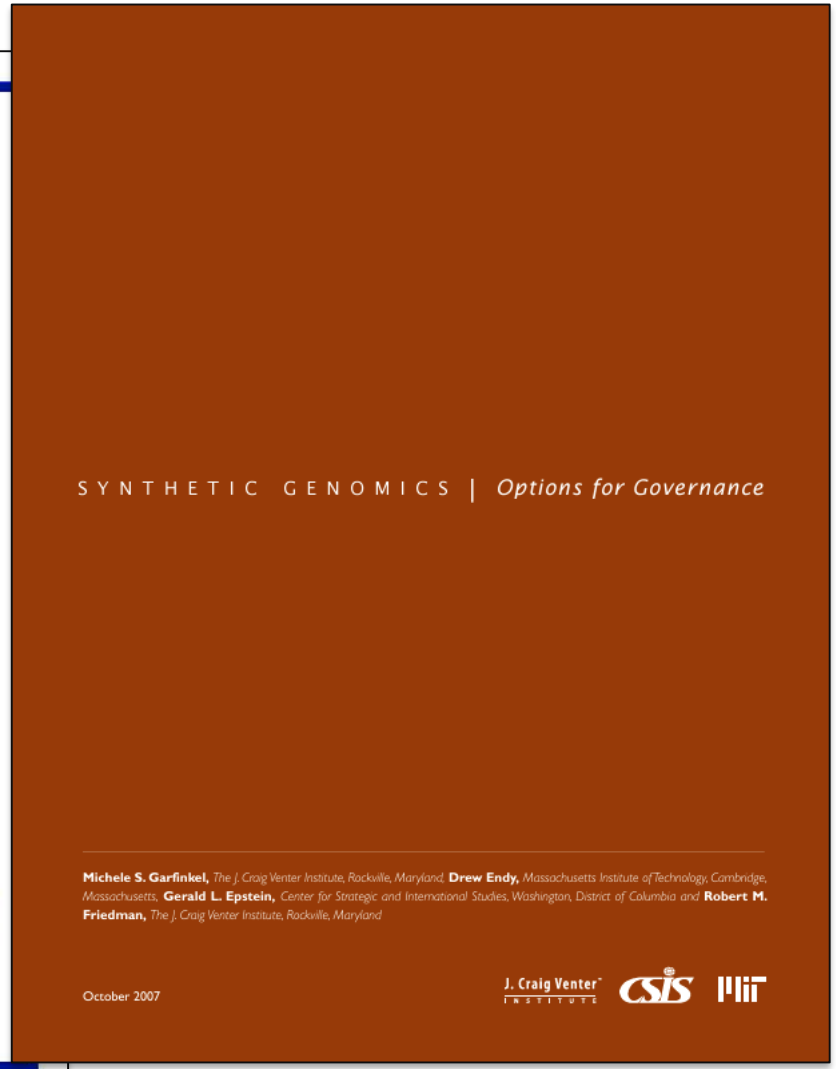
2003



2006



2007



2006: The Guardian buys variola

The Guardian US

This article is more than **18 years old**

Did anyone order smallpox?

But ordering part of this long-dead pathogen's DNA proved easier than anyone dared imagine. All it took was a invented company name, a mobile phone number, a free email address and a house in north London to receive the order by post.

Last November New Scientist magazine surveyed 12 gene synthesis companies in North America and Europe. Only five said they always screened their orders for suspect sequences, and three said they never did. These were all doing relatively large-scale synthesis, providing sequences a few hundred letters long, but there are many more companies such as VH Bio Ltd which make so-called oligonucleotides, sequences around 100 letters or smaller.

2009: Formation of the IGSC

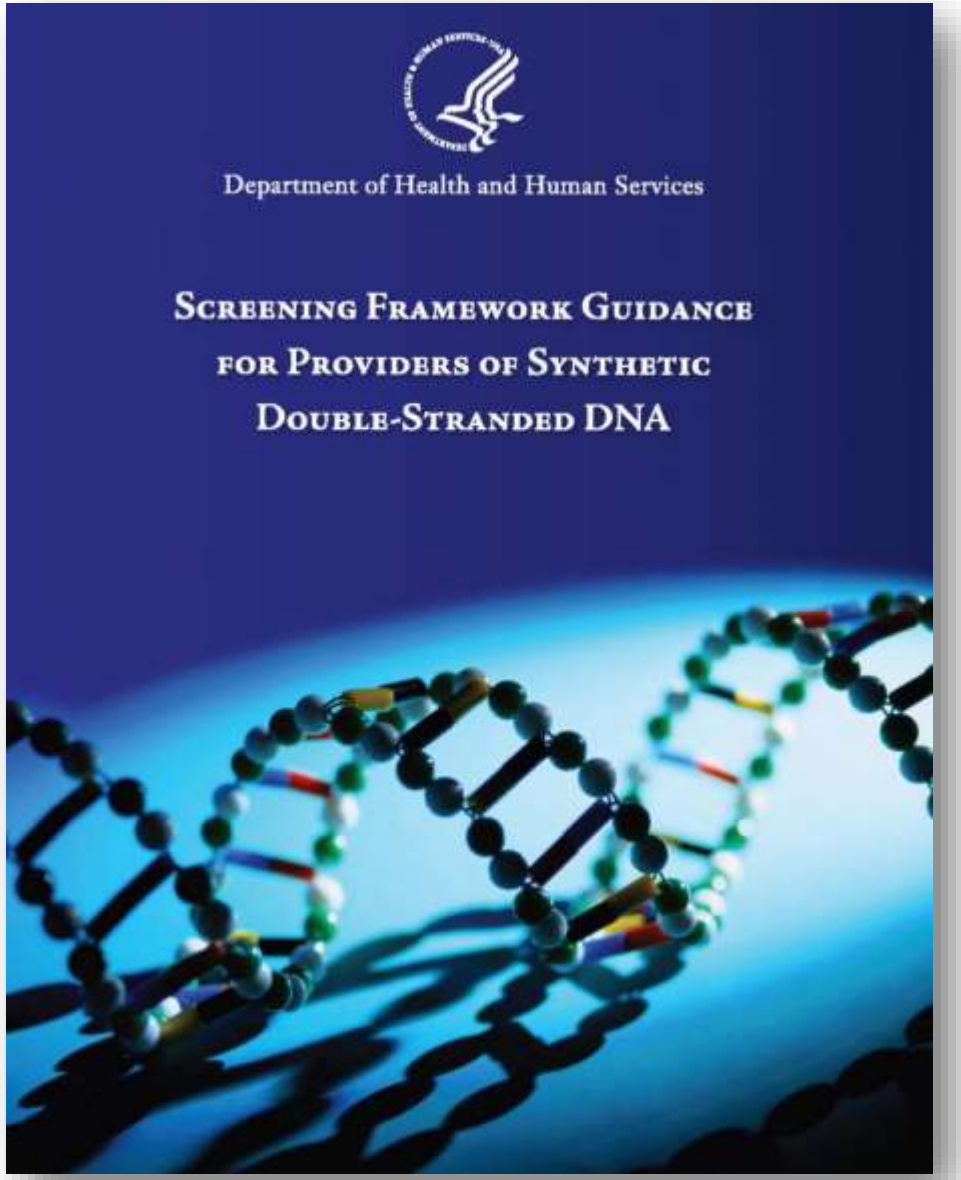
- 2007-2009, many engagements between industry, government and civil society groups
- **Industry and government concerns were aligned**: no one wanted misuse!

- 5 of the largest global synthesis companies formed the IGSC to coordinate development of best practices

2010: US releases Screening Framework Guidance

- US government responded to industry requests for guidance

- Guidance recommended (but did not require):
  - Screen customers for legitimate purpose
  - Screen sequences for ‘concern’
  - Alert the FBI if an order is suspicious
  - Retain all records for at least 8 years



~15 years go by …

- Technology for DNA synthesis advances
- Costs come down, throughput goes up
- Several start ups work on benchtop synthesis
- US IARPA runs the FunGCAT program
- The IGSC grows to 40+ member companies
- *Continued engagement with governments*

IGSC
international gene synthesis consortium™

International Gene Synthesis Consortium

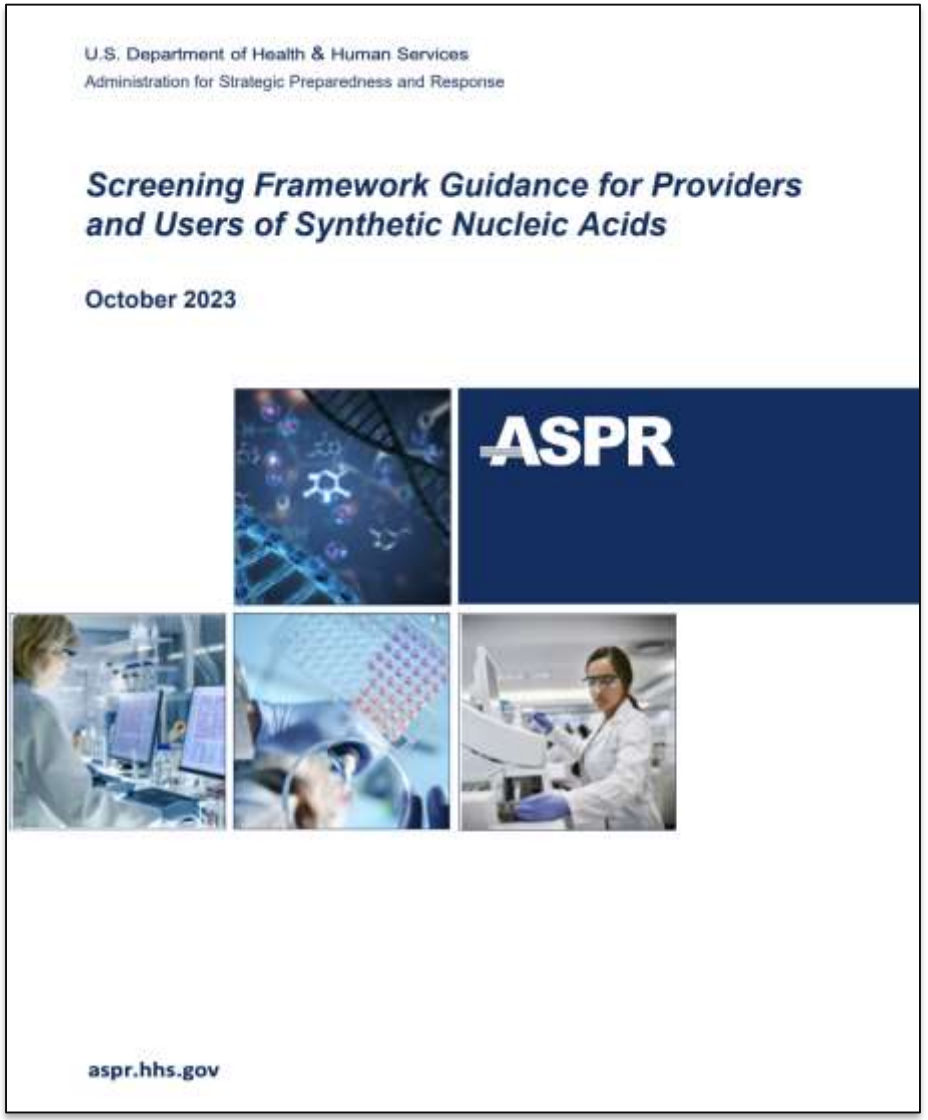
TWIST BIOSCIENCE

Where are we now?

2023: US releases Updated Guidance

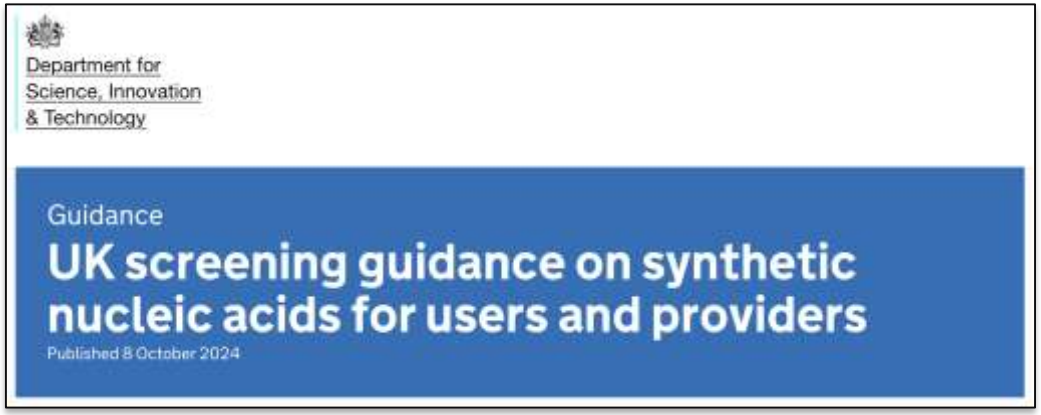
- Produced after consultation with industry and with extensive inter-agency USG collaboration

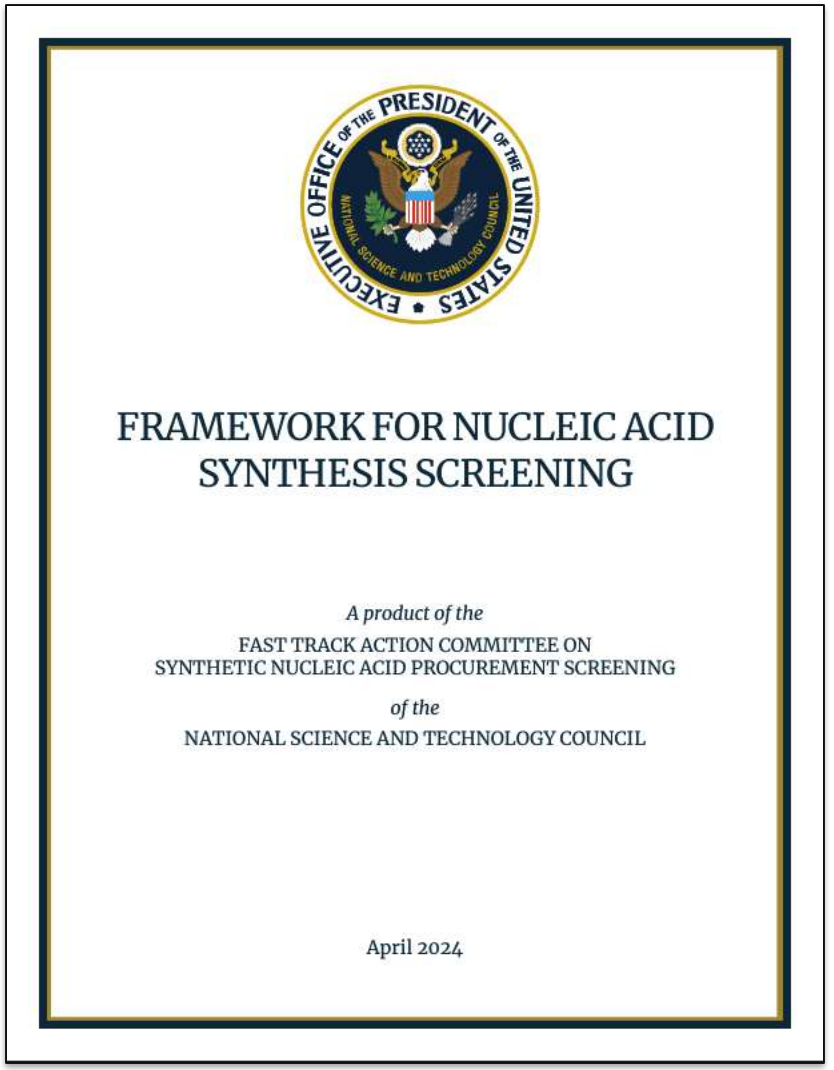
- Incorporated biosecurity guidance for benchtop synthesizers
- Expanded to cover DNA and RNA, ss or ds
- Reduced sequence window from 200bp to 50bp
- Encouraged industry to develop a wider definition of ‘sequence of concern’
- **Did not address DNA data storage**



2024: US releases a Framework, UK releases guidance

- Requires producers of synthetic DNA (including benchtop manufacturers) do six things:
  - Attest to implementing the Framework
  - Screen sequences from orders
  - Screen ordering customers
  - Report suspicious orders to law enforcement
  - Retain purchase order and screening records
  - Ensure cyber/information security
- Limits federal R&D dollar spending on synthetic nucleic acids only to companies that attest

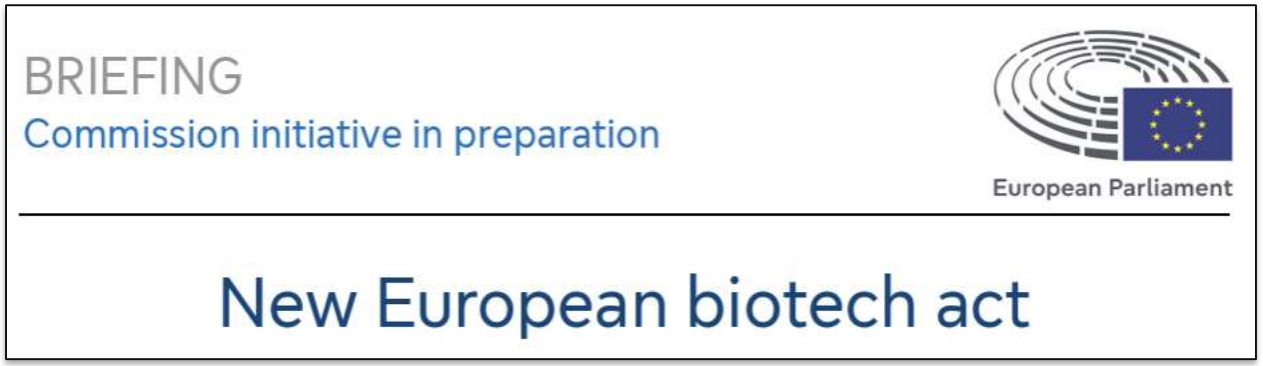




2025: US rewriting its framework/EU drafting its Biotech Act

- U.S. is rewriting its 2024 framework
  - May still limit federal funds to companies with biosecurity practices
  - Might introduce third-party verification for the first time
- EU Biotech Act
  - Considering possible requirements around biosecurity screening for synthetic nucleic acids

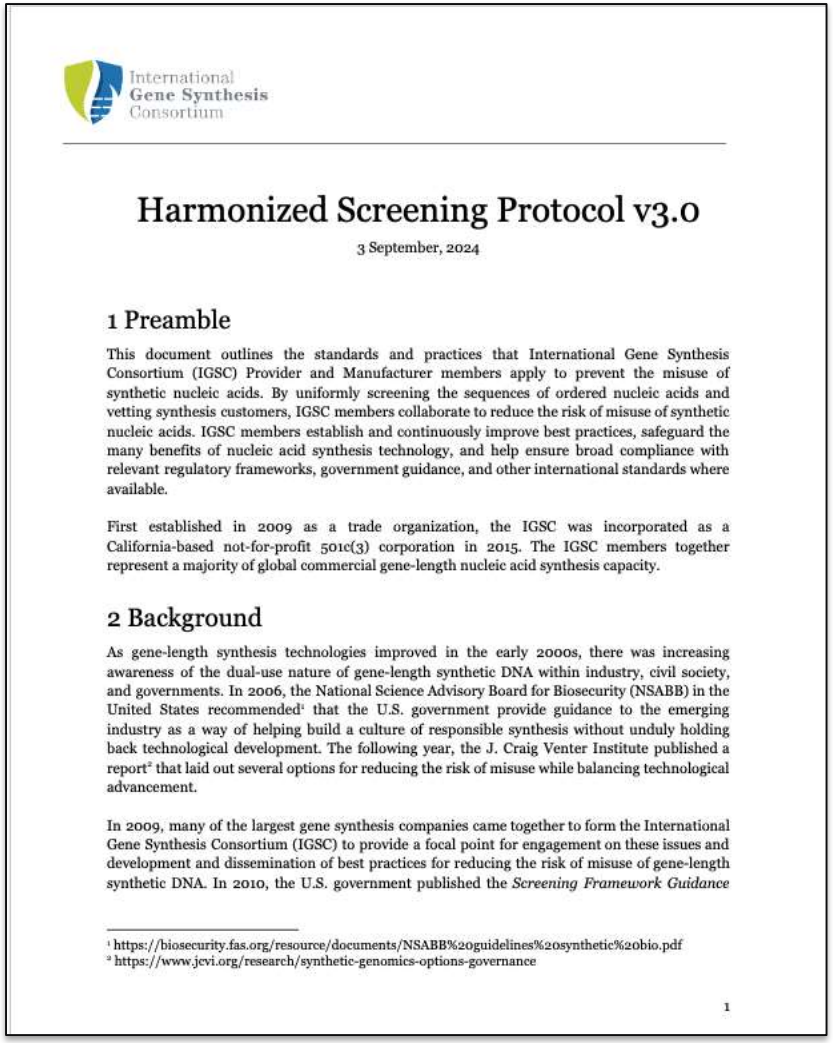
- **AI design tools opening new questions around sequence screening**



IGSC today

- 40 members from 10 countries
- Monthly meetings to share news, best practices or concerning experiences

- IGSC products:
  - Harmonized Screening Protocol v3.0
  - Restricted Pathogen Database (RPD)
  - Suspicious Order Reporting Form
  - Test Set Working Group



International Gene Synthesis Consortium

Harmonized Screening Protocol v3.0

3 September, 2024

1 Preamble

This document outlines the standards and practices that International Gene Synthesis Consortium (IGSC) Provider and Manufacturer members apply to prevent the misuse of synthetic nucleic acids. By uniformly screening the sequences of ordered nucleic acids and vetting synthesis customers, IGSC members collaborate to reduce the risk of misuse of synthetic nucleic acids. IGSC members establish and continuously improve best practices, safeguard the many benefits of nucleic acid synthesis technology, and help ensure broad compliance with relevant regulatory frameworks, government guidance, and other international standards where available.

First established in 2009 as a trade organization, the IGSC was incorporated as a California-based not-for-profit 501c(3) corporation in 2015. The IGSC members together represent a majority of global commercial gene-length nucleic acid synthesis capacity.

2 Background

As gene-length synthesis technologies improved in the early 2000s, there was increasing awareness of the dual-use nature of gene-length synthetic DNA within industry, civil society, and governments. In 2006, the National Science Advisory Board for Biosecurity (NSABB) in the United States recommended[1] that the U.S. government provide guidance to the emerging industry as a way of helping build a culture of responsible synthesis without unduly holding back technological development. The following year, the J. Craig Venter Institute published a report[2] that laid out several options for reducing the risk of misuse while balancing technological advancement.

In 2009, many of the largest gene synthesis companies came together to form the International Gene Synthesis Consortium (IGSC) to provide a focal point for engagement on these issues and development and dissemination of best practices for reducing the risk of misuse of gene-length synthetic DNA. In 2010, the U.S. government published the *Screening Framework Guidance*

[1] https://biosecurity.fas.org/resource/documents/NSABB%20guidelines%20synthetic%20bio.pdf
[2] https://www.jcvi.org/research/synthetic-genomics-options-governance

1

Biosecurity at Twist

Upon submission to Twist, all gene synthesis orders are subject to:

- **Customer screening**
  - Verifies legitimacy with respect to an institution and identity with respect to an individual
  - Is the customer or institution on any lists maintained by the Departments of Commerce, State and Treasury?
  - Is the customer licensed to carry out work on regulated pathogens?
  - Ensure the shipping address is not a PO Box or private residence

- **Sequence screening**
  - Does the sequence pose a significant dual-use **or biosafety risk**?
  - Is the sequence a ’best match’ to a regulated bacterial or viral pathogen?
  - Is the sequence legal to manufacture and ship within the United States?
  - Is the sequence from a gene that can endow or enhance pathogenicity?
  - Does the sequence require an export license to ship overseas?

Biosecurity for DNA Data Storage

Community-led risk reduction for emerging technologies

- Continue the ~20 year history of community-led risk reduction
- **Threat model:** an actor presents a pool as ‘for DNA data storage’ but a minority of the sequences in the pool can be recovered and used to build a viral genome.
  - E.g. 200 oligos in a 1B oligo pool

- Sequence screening is already computationally expensive at *gene sequence scales*
- It will be far too expensive (vs. market price of pool) to screen e.g. 60B oligo pools

- There are some computationally efficient screening methods!
  - Eval of hits is complicated – what is incidental homology vs. intended acquisition?
  - Need a PhD to sit and make those decisions? At what scale? How fast?

Data Storage – Trusted Device

Treat the device/actor as a black box – binary data in, DNA out

- No opportunity for an untrusted third party to intervene & add sequence
- Sequences may have *incidental homology* to pathogens
- Acquisition of pathogens via this route is extremely difficult

- This is the ***easiest*** route in the current regulatory & political environment
  - Specific exemptions in future legislation or regulatory language for devices that encode+synthesize

Data Storage – Trusted Storage

Treat the pool as untrusted but the custodian as trusted

- Assume pool definition is ‘untrusted’ – might contain pathogen DNA
- Synthesis provider stores the encapsulated pool without allowing physical access to the pool
- Customer can request (binary) data recovery at any time

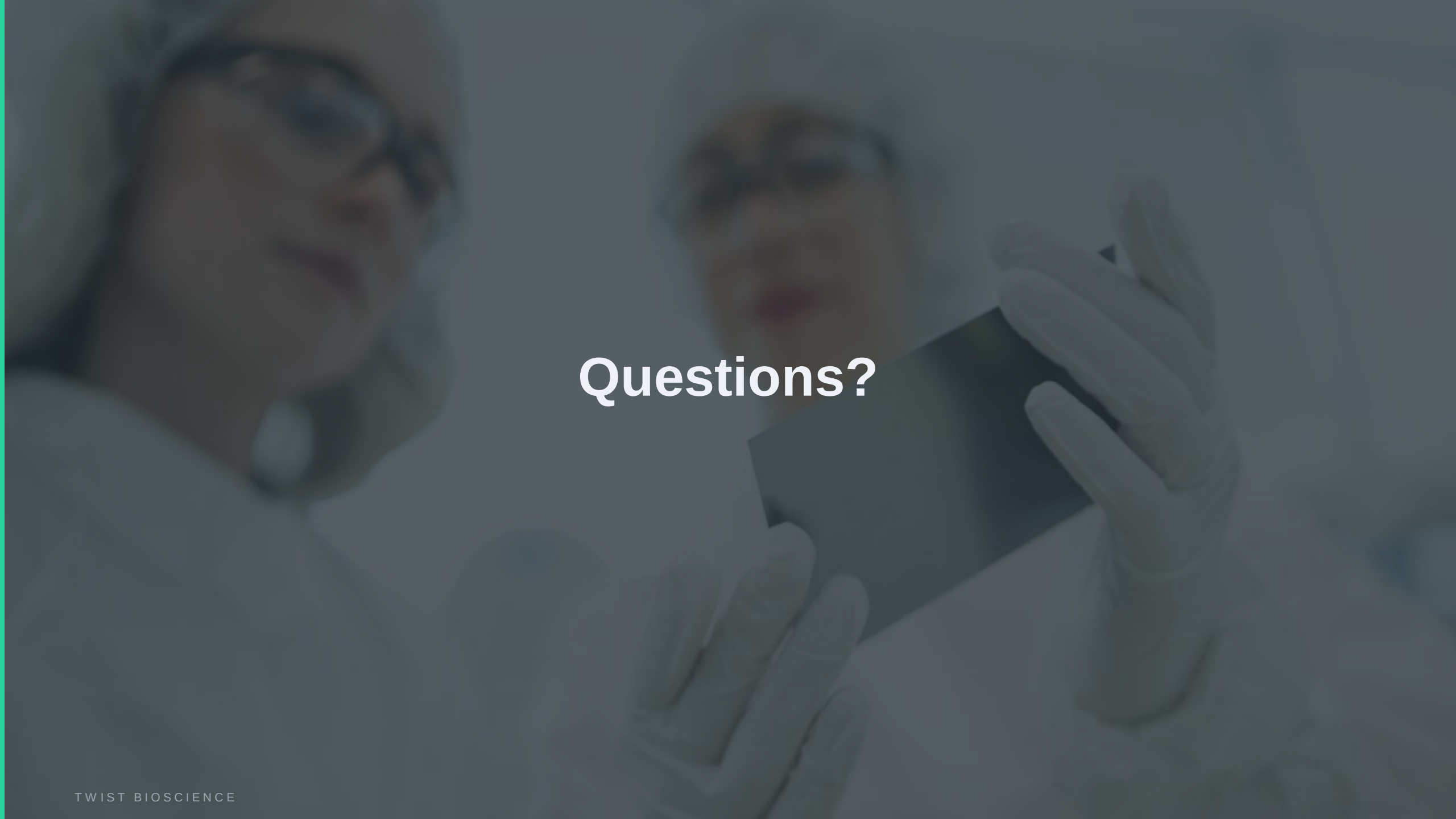
- This is **slightly more complicated** – requires:
  - synthesis provider to also be a physical storage and recovery provider
  - trust on the part of the customer to relinquish physical control of their stored data

Data Storage – Trusted Parties

Build an ecosystem of trusted parties

- Separating encoding from synthesis creates *uncertainty* of pool security
  - MITM ‘attack’ could add oligos to an encoded pool before synthesis
- Lean in to US/UK government recommendation around *customer* screening
  - Only work with ‘legitimate’ companies/actors
- Could eventually mirror something like cert chains for distributed HTTPS trust

- This is **more complicated** – requires:
  - political education and acceptance
  - technological implementation (of trust chains)
  - education around Restricted Party Screening best practices


Questions?