

Three Data Building Blocks to a Better NSE

Camille Mathieu
Knowledge Management

Strategic Deterrence
Lawrence Livermore National Laboratory

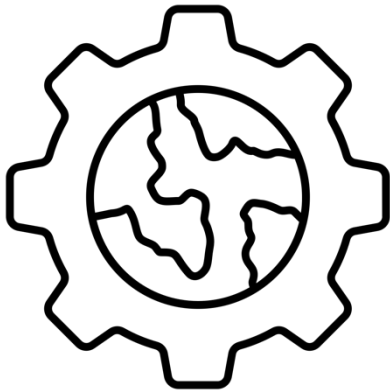
DOE Data Days

October 2024



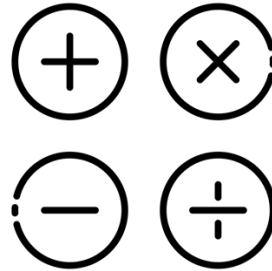
Bottom-Line Up Front

- A transformed, data-driven NSE requires:



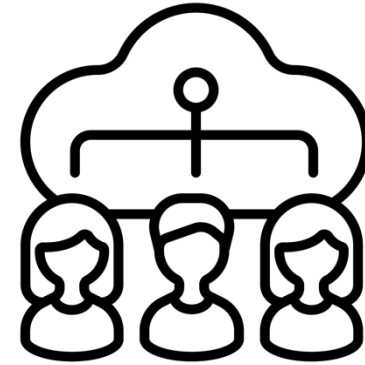
Block 1

A defined environment that enables discrete functions



Block 2

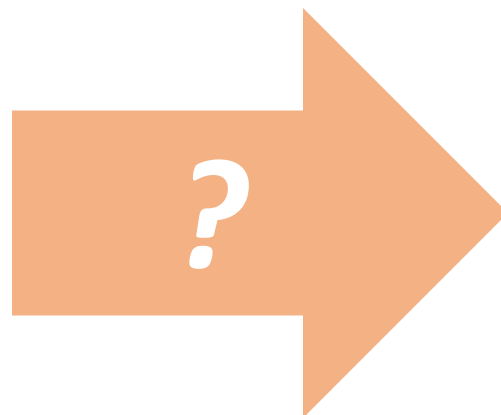
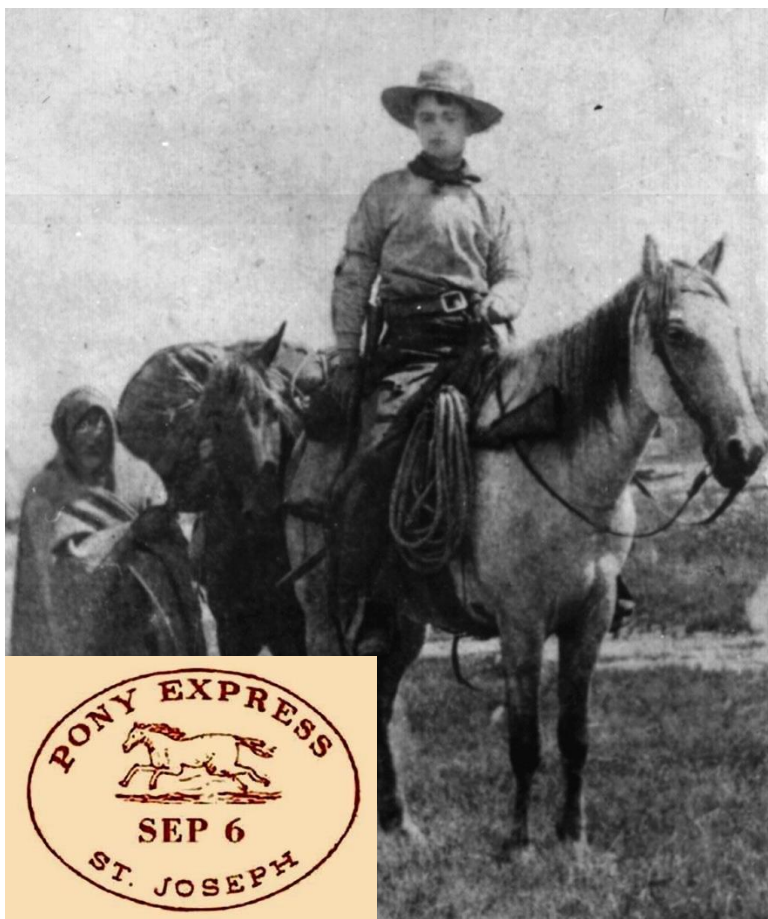
Normalized inputs into the environment

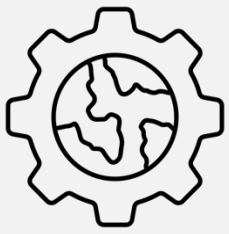


Block 3

Participants who can perform discrete functions

How do you get from the Pony Express to today's USPS?





Block 1

A defined environment that enables discrete functions

The USPS requires the following *environment*:

- Named states and cities
- Numbered buildings
- Cross-state transportation
- Funding mechanism

To enable the following discrete *functions*:

- ✓ Direct-to-home delivery of parcels
- ✓ Pickup of outgoing mail
- ✓ Returning undeliverable mail to sender



Block 1

A defined environment that enables discrete functions

**The NSE requires the following
*environment:***

- ? ESN Hub
- ? Local lab SRD networks
- ? Classified email
- ? PDM Link, other repos

**To enable the following discrete
*functions:***

- ? Sending CAD files between sites
- ? Model V&V
- ? Sharing documents & data



Block 1

A defined environment that enables discrete functions

The NSE requires the following

To enable the following discrete

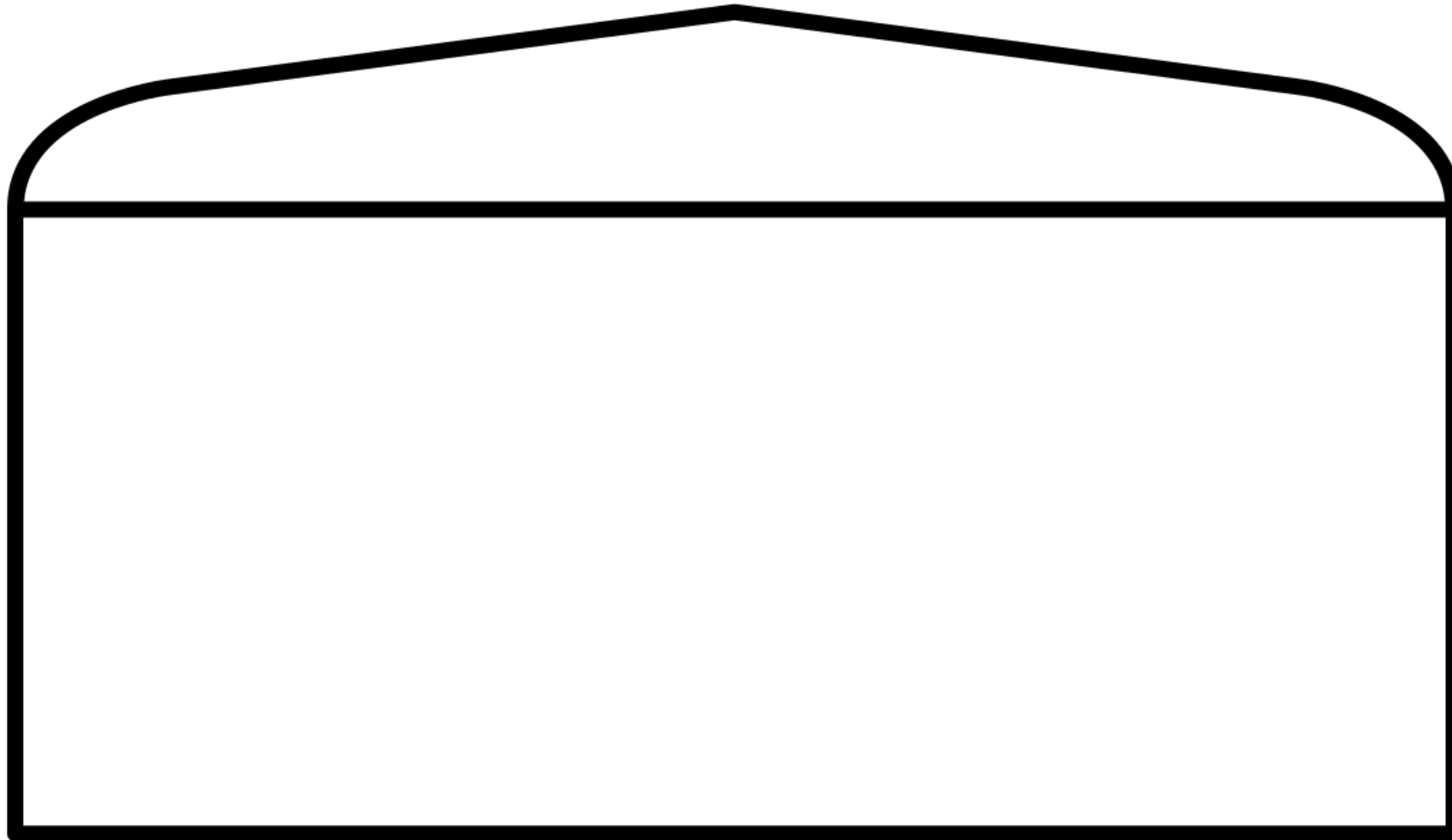
NSE cannot make progress on digital engineering goals until we can
1) constrain the environment we expect to digitally-engineer in and
2) enumerate the specific functions this environment will enable.



Block 2



Normalized inputs into the environment



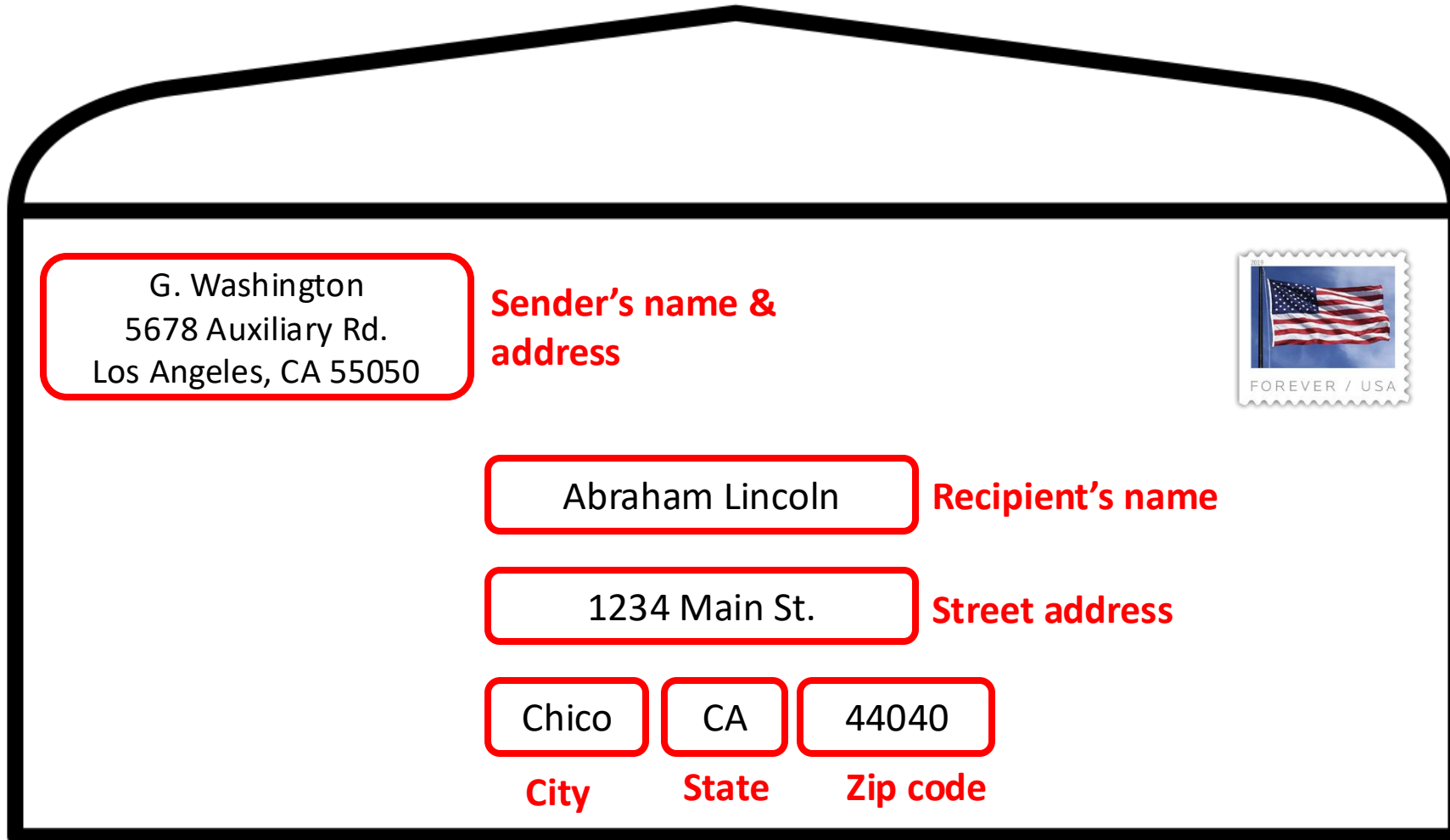
*How do you
address an
envelope?*



Block 2



Example of normalized inputs



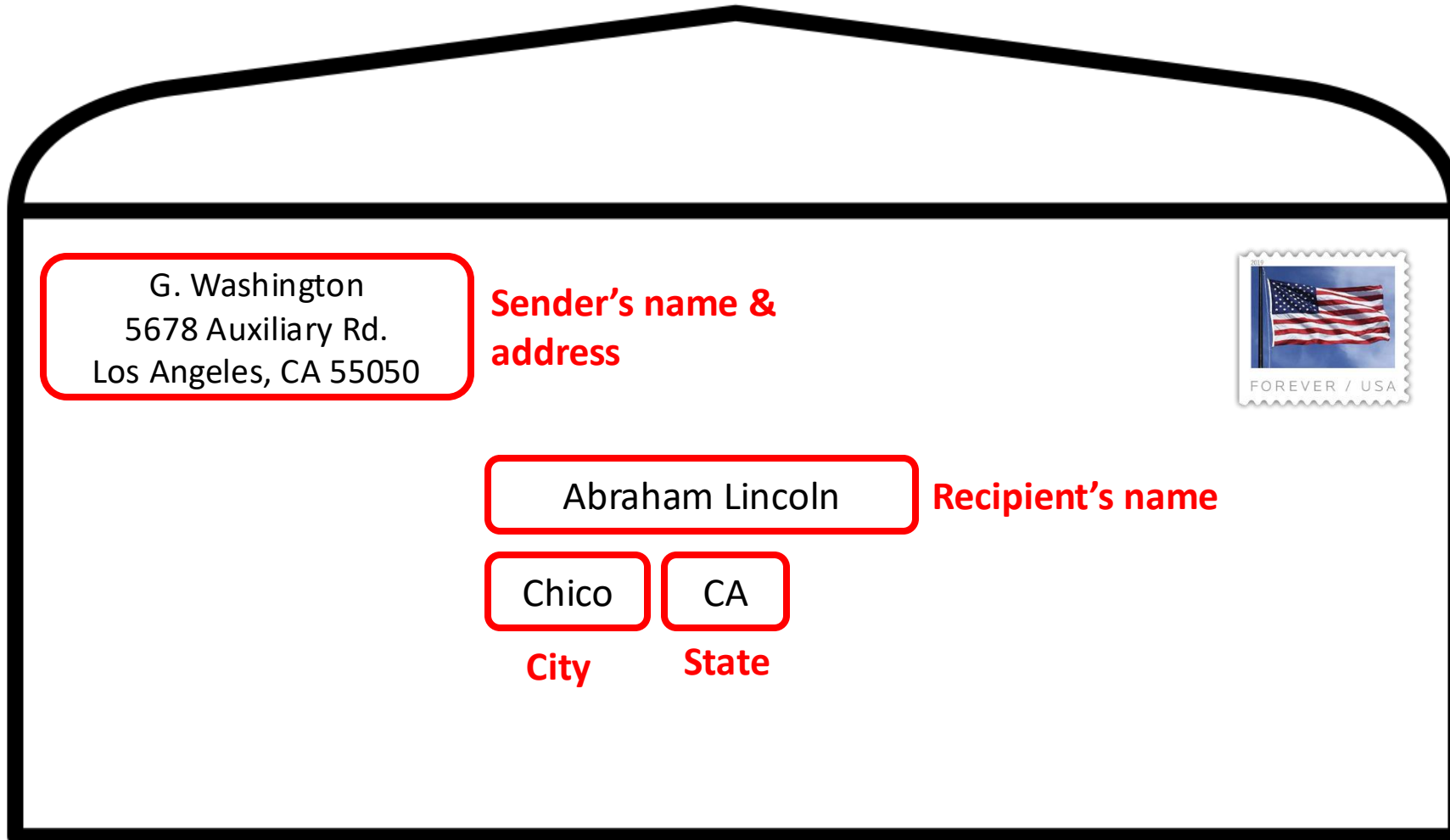
How do you address an envelope?



Block 2



Example of normalized inputs: Field standards



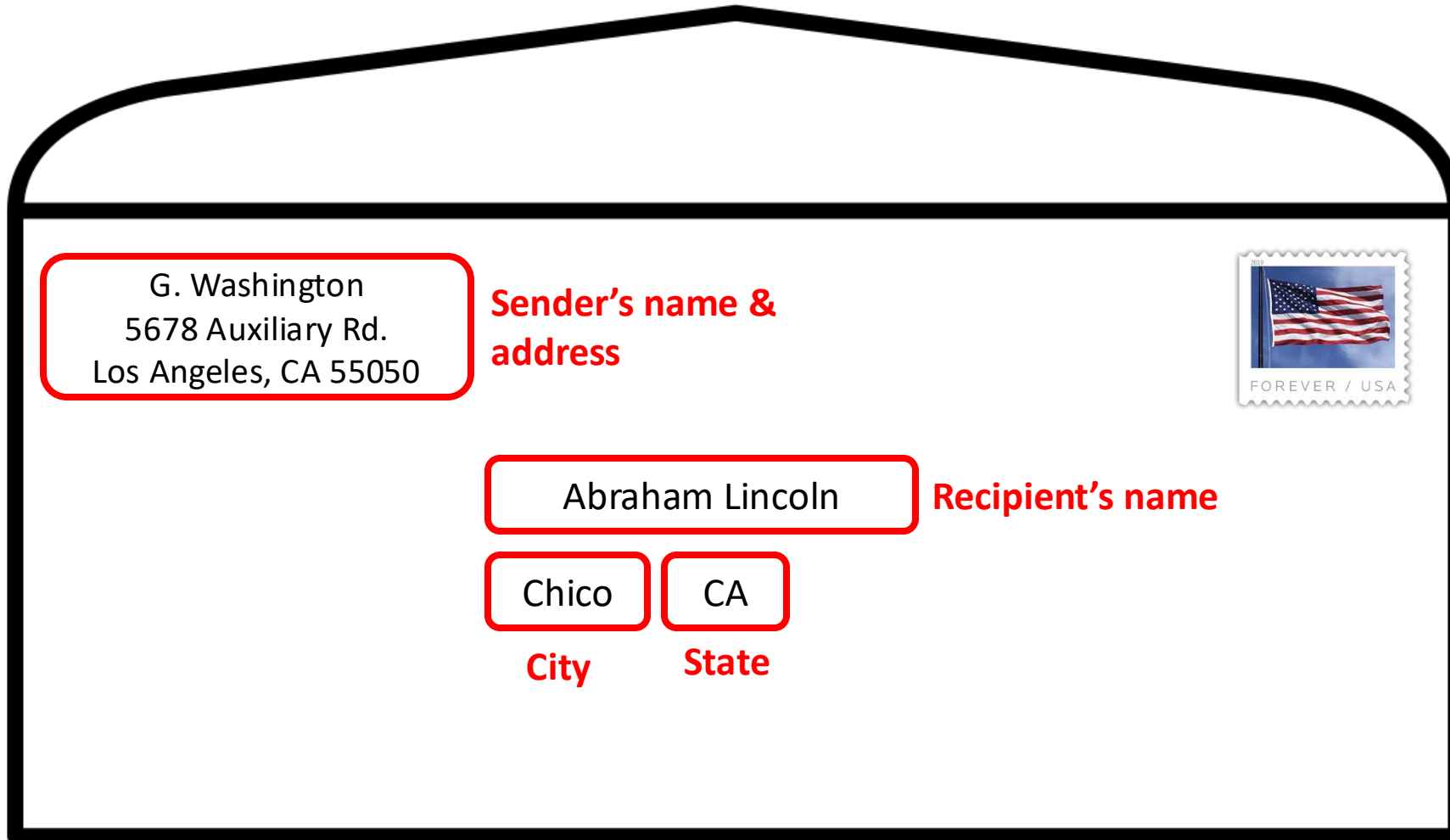
Why is this letter undeliverable?



Block 2



Example of normalized inputs: Field standards



Why is this letter undeliverable?

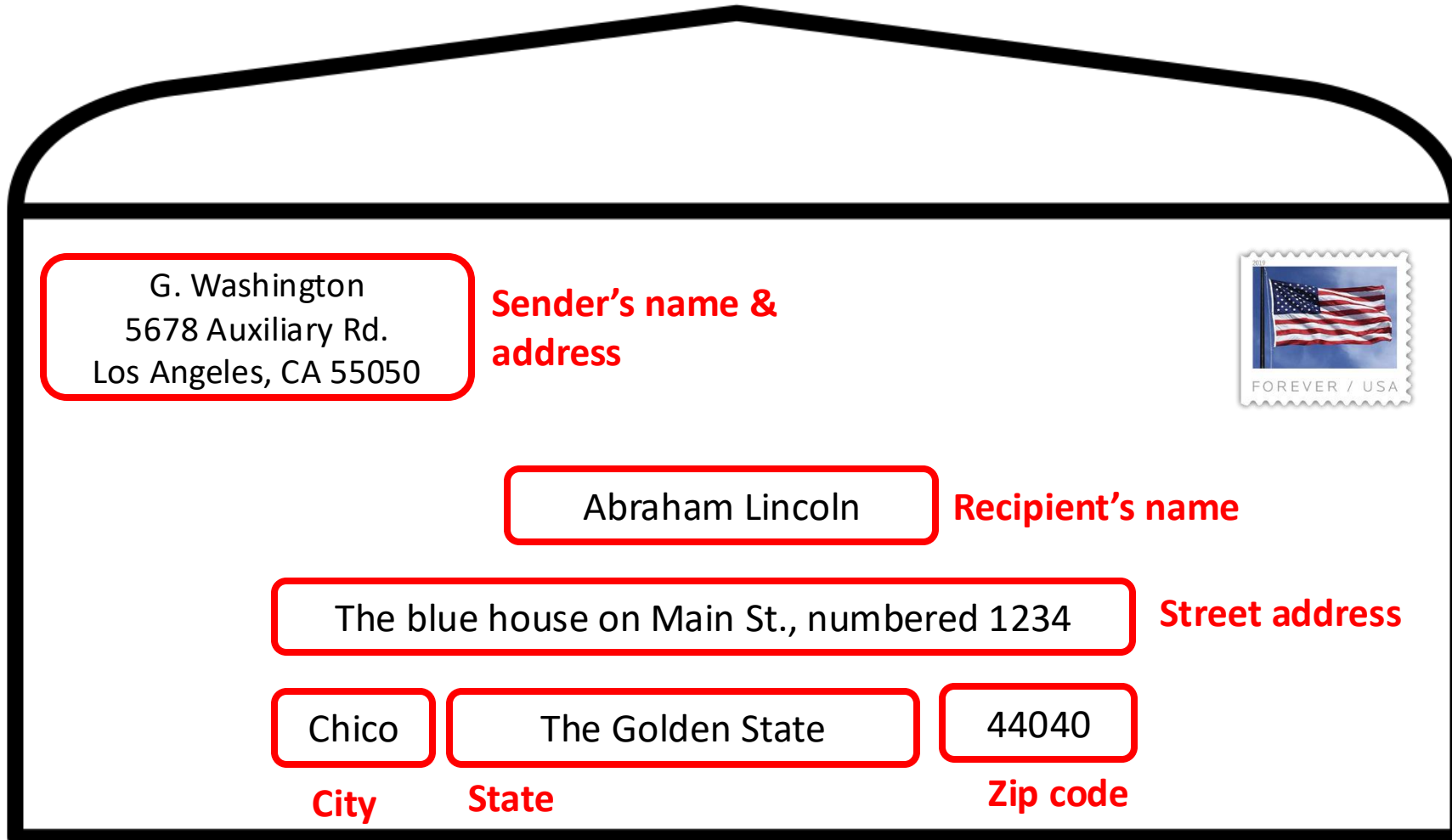
Because not all the standard *fields* required for USPS processing are included.



Block 2



Example of normalized inputs: Value standards



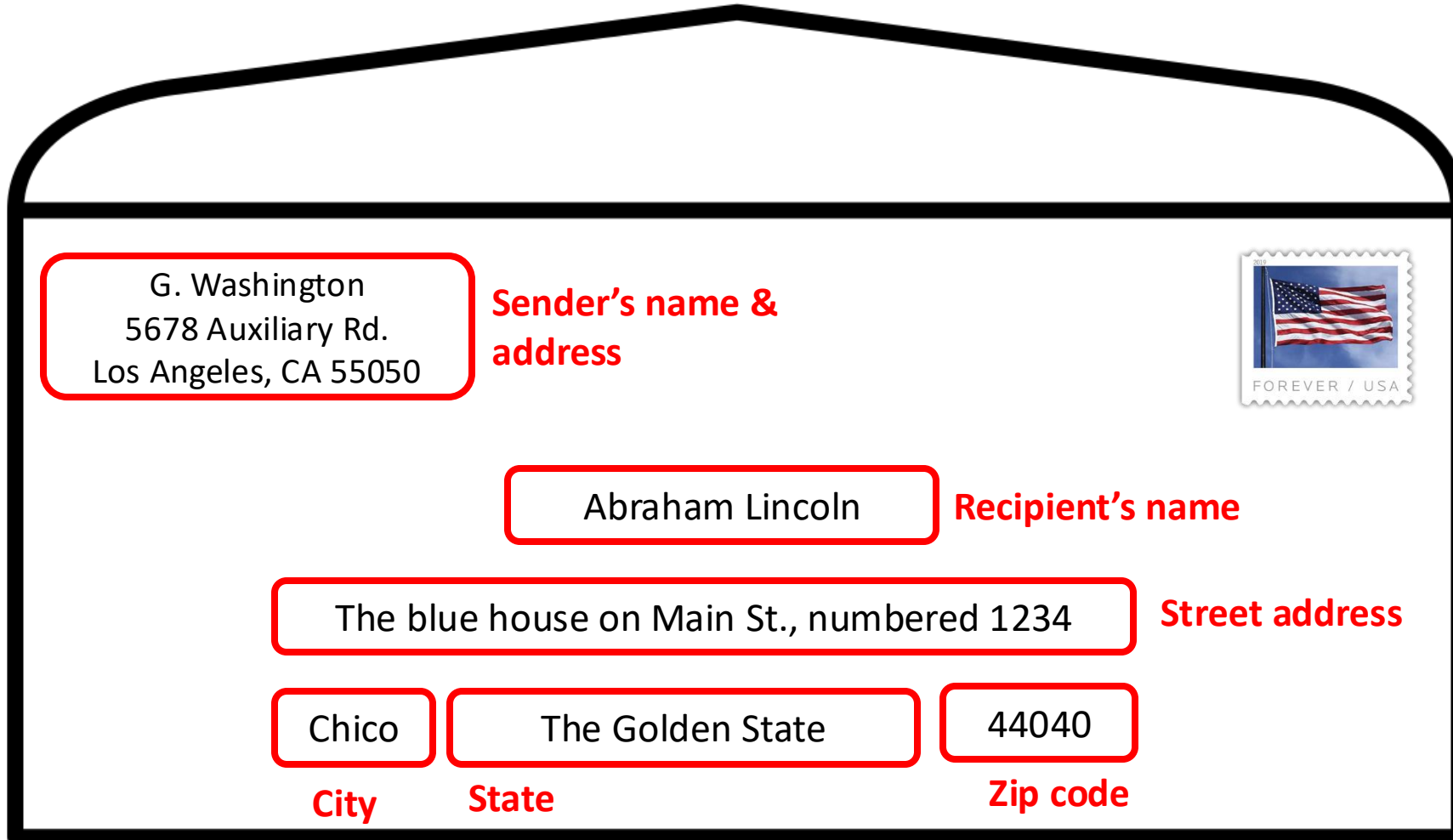
Why is this letter undeliverable?



Block 2



Example of normalized inputs: Value standards



Why is this letter undeliverable?

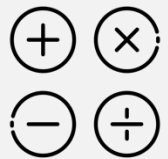
Because the information, while not incorrect, does not conform to USPS *value* standards

⊕ ⊗ Block 2

⊖ ⊕ Field & value metadata standards

Field standards	What information a data object must be able to tell us about itself	→ <i>nse:InitialTemperature</i>
Value standards	What language it must use to be able to tell us that information	→ <numerical value in Kelvin>

- USPS has *Mailing Standards of the United States Postal Service Publication 28 - Postal Addressing Standards*
- What should the NSE version of *Publication 28* say?



Block 2

Field standard example


DRAFT

Metadata Working Group (MWG)

Metadata Standards for the NSE: Core Fields

Recommendation to the National Security Enterprise (NSE) for the
standardization of core metadata fields in support of improved inter-lab
information sharing and advanced digital engineering

December 2024

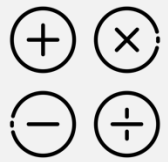


NNSA
National Nuclear Security Administration

Prepared by:
Los Alamos National Laboratory
National Security Research Center, Weapons Research Services

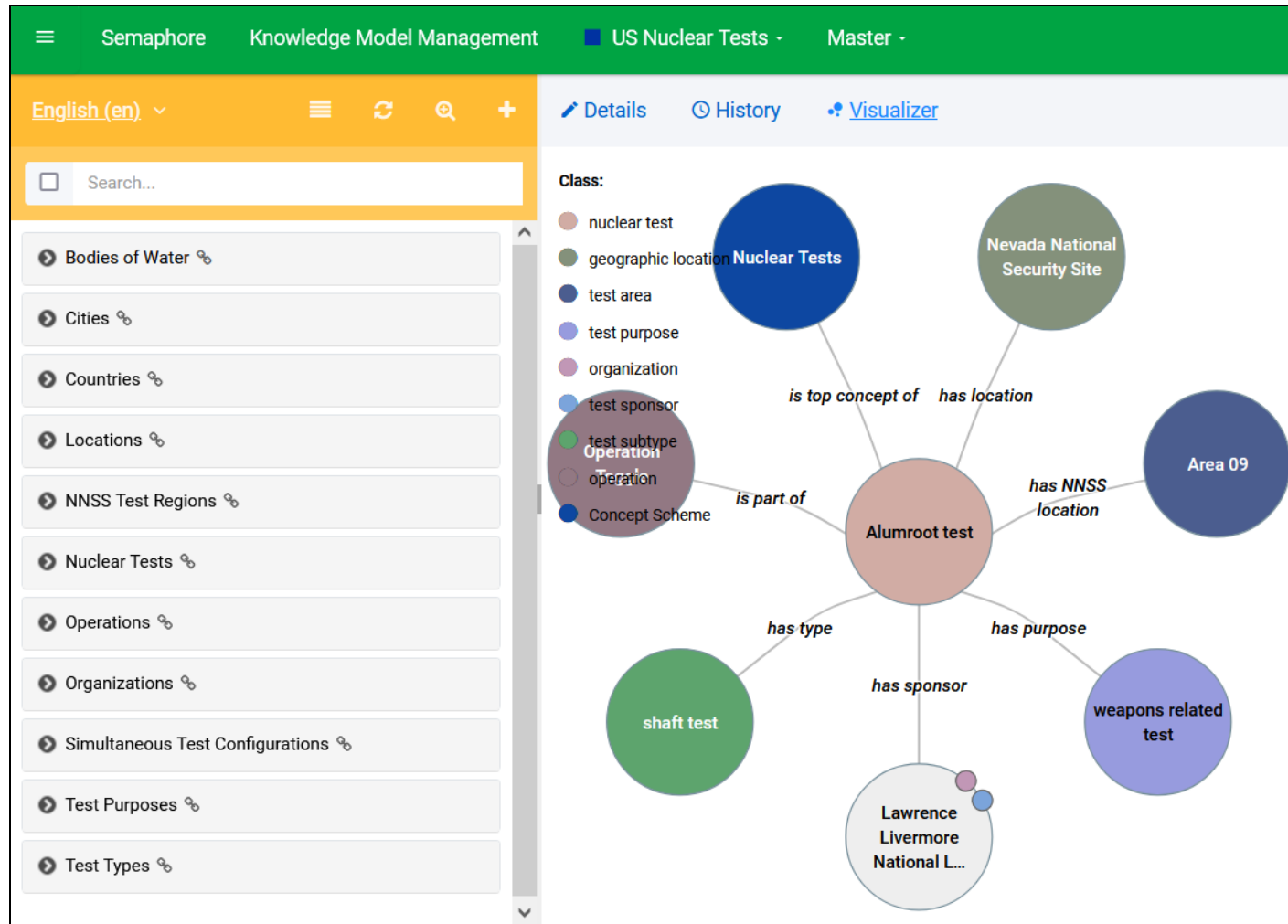
Lawrence Livermore National Laboratory
Strategic Deterrence Knowledge Management Program

- Ongoing LLNL/LANL collaboration to recommend the NSE use the following 6/7 fields as “core”:
 - Title (*nse:Title*)
 - Identifier (*nse:Identifier*)
 - Date (*nse:Date*)
 - Contributor (*nse:Contributor*)
 - Originating Organization (*nse:OriginatingOrganization*)
 - Sensitivity (*nse:Sensitivity*)
 - Caveat (*nse:Caveat*, SRD only)
- These are core descriptive fields only; future standards may make specifications on data representation fields (like *nse:InitialTemperature*)

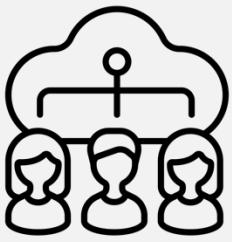


Block 2

Value standard example



- LLNL uses a centralized Ontology Management System (OMS) to capture & distribute values
- These resources are in a standard format to be shared across the NSE
 - Ex: list of all US nuclear tests based on NV-209 is available for download on ESN Wiki


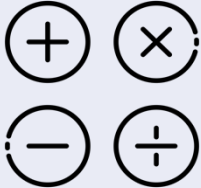



Block 3

Participants who can perform discrete functions

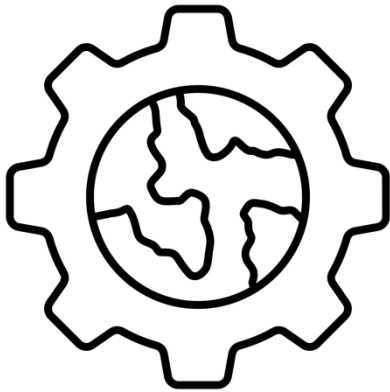
- In order to operate efficiently in this ecosystem, people need to know how best to submit, receive, and manage their content
 - Need to know the **rules of the environment**
- We as the info management people need to define the optimal operation of the system overall. Consider:
 - What percentage of your lab's work is emailed out once, then put in a shared drive?
 - Do people know when to use PDM Link vs Granta vs Online Vault?
 - What is the function of SharePoint in your environment? The function of Confluence? When would you recommend one over the other?
 - How often are new employees waiting weeks/months to get added to the necessary access groups?
- How can we expect employees to operate a digital-first enterprise if no one can define the “best use” of the environment end-to-end?

Benefits of investing in these blocks

If we have...	Then we can expect...
 <p>Block 1 A defined environment that enables discrete functions</p>	<ul style="list-style-type: none"> • Digital engineering foundations to be established and usable • Consistent platform for tracking digital thread
 <p>Block 2 Normalized inputs into the environment</p>	<ul style="list-style-type: none"> • Easier sharing and reuse, both across sites and within local labs • Long-term context for information search and retrieval • To produce data that is born AI-ready • Value standards expressed as semantic resources (ontology/taxonomy) can be used in RAG applications
 <p>Block 3 Participants who can perform discrete functions</p>	<ul style="list-style-type: none"> • Shorter time to competency for new staff • Information will not be lost by exclusion from source systems

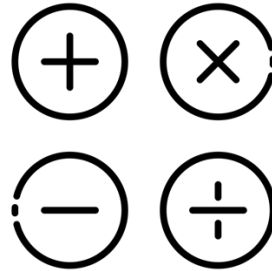
Bottom-Line Up Front

- A transformed, data-driven NSE requires:



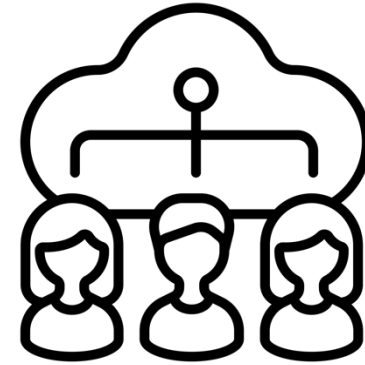
Block 1

A defined environment that enables discrete functions



Block 2

Normalized inputs into the environment



Block 3

Participants who can perform discrete functions

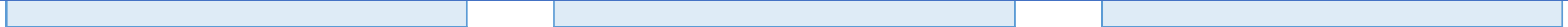
Bottom-Line Up Front

- A transformed, data-driven NSE requires:



Takeaways

- Local efforts should be exposed and collaborated on so that we can each do a part of the work without duplicating each other
- Ultimately, success will require that elements of each block are written into requirements and passed down from authoritative bodies





Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.