

APPENDIX TABLE 5A-3. Dimensions of data challenges in the pilot year’s five topic areas.

| | Situational awareness | AI |
|------------------------------|---|---|
| Type of data | Patents/publications/documents/citations/funding/production facilities/supply chains | Use of AI in firms/progress of AI/patents/citations/employment data |
| Ease of access | <p>Relatively easy access to publicly available publications/patents, etc., but requires extensive curation and can require expensive licensing. Skewed toward Western and English-speaking countries. Less access to funding data at a granular level. Not possible to identify data that can be compiled in an intersectional way. Patents not representative of full body of inventive activities, because of trade secrets.</p> <p>Limited or no access to product, production, and supply chain data, which are mostly held by private firms</p> | Unclear what to track/standards of measurement |
| Accuracy/completeness | Limited to scope of data sources and languages of publication or countries of patenting. May be missing researchers and institutions, lacking comparability of documents or technology descriptions, unpublished work, and work in other languages | Different datasets have different limitations. Job postings are limited to those that post on that site. Patents only cover a percentage of activities. Surveys limited by response quality, rate, and population. Census surveys can be mandatory. |
| Timeliness/frequency | Publication speed, some ability to see in real-time with preprints, technical reports, or venues like the Social Science Research Network | Surveys take a longer time but are more accurate. Patents take 2–3 years to come out. Job postings are immediate. |
| Cost of validating | <p>Low cost: peer-reviewed papers</p> <p>Higher cost: preprints, technical reports, or venues like the Social Science Research Network</p> | High |
| Data suppression | Corporate control of publications; governmental control of publications; privacy concerns with sociodemographic data | Incentives not to patent. Depending on the company, some incentives not to publish. Top-caliber individuals may not be recruited through sites. |

| | Semiconductors | Biopharmaceuticals | Energy and critical materials |
|-----------------------|---|---|---|
| Type of data | Production data/trade data | Production data/clinical trials/trade data/FDA | Production data/trade data |
| Ease of access | Difficult to access from different firms, linkages aren't available/accessible, treated as parts of other products (aggregation issues) | FDA data publicly available (easy to access), granular production in other countries difficult | Expensive to access production data from aggregators (e.g., Bloomberg) |
| Accuracy/completeness | Lack of linkages, unclear production sites (what is produced), missing data, aggregated at high level; treated as parts of other things (aggregation issues) | Unclear production sites, differences in different types of data (e.g., devices vs. pharma), level of aggregation (especially for inputs) | Missing trade and production data, aggregation of different chemical products, uncertainty around types of intermediary inputs used |
| Timeliness/frequency | Production data: firm-level (yearly) Trade data: country-level (monthly) | Production data: firm-level (yearly) Trade data: country-level (monthly) | Production data: firm-level (yearly) Trade data: country-level (monthly) |
| Cost of validating | Extremely high cost/potentially impossible to validate (production data) Trade data easier to validate but potential issues around noise; difficulty of auditing foreign manufacturers | Lower cost for FDA-approved devices/institutions; difficulty of auditing foreign manufacturers | Extremely high cost, potentially impossible to validate at individual level (production data); possible to validate by end product sales. Trade data easier to validate but potential issues around noise and missing trade; difficulty of auditing foreign manufacturers |
| Data suppression | Incomplete disclosure for competitiveness reasons | | Incentives exist for foreign/illicit entities to hide/reroute trade to avoid tariffs/sanctions/embargoes |