International Nucleotide Sequence Database Collaboration (INSDC):

Membership Acceptance and Performance Guidelines

Introduction

INSDC Members are expected to adhere to the criteria specified in the INSDC Membership Acceptance and Performance guidelines (guidelines) as defined here. Members are expected to participate in routine review to assure their continued adherence. These guidelines include a resource intended for use by prospective Members as a self-assessment of the technical and organizational capabilities in place within their institution and their alignment to the capabilities required of Full and Associate Members (see Appendix) for participation in sequence data sharing across INSDC.

Assessment Areas

Prospective INSDC Members must demonstrate readiness in the following five areas:

- Governance and institutional context
- Technical infrastructure
- Data operations
- Communications and engagement
- Quality assurance

Associate Members must demonstrate maturity in each of these areas sufficient to collect, archive, and present a subset of sequence data for search, analysis, and retrieval that is limited in scope either by region, scientific domain, and/or by data type. Full Members must demonstrate the maturity to scale these activities to handle all nucleotide sequence data types in scope for INSDC as defined by the INSDC Executive Committee (EC).

Each of the five areas is subdivided into specific categories with indicators to be used for self-assessment of maturity. Below is a description of each of the above categories and rationale for inclusion in the guidelines.

Governance and institutional context assesses an institution's ability to strategically plan, manage, and deliver data resources at the scale required by INSDC. It includes institutional *governance* processes, *vision and strategy*, and *sustainability*. Governance refers to whether institutions have transparent processes that are independent and free to make decisions for the benefit of global science. Vision and strategy refers to future planning and ambition in terms of sourcing, collecting, and sharing data globally. Sustainability refers to the plans and funding needed for continuity of service, disaster recovery, and succession planning.

Technical infrastructure evaluates performance aspects of the technology that will support INSDC activities. It includes assessment of the *software stack*, *storage systems*, *network* and *computing infrastructures* that the institution has in place. In general, the categories in this area enable a prospective Member's self-assessment of institutional ability to robustly and reliably scale to meet the demands of participation in INSDC. In this area, the demonstration of technical infrastructure maturity is highly dependent on the desired membership level. Full INSDC membership, which involves handling of all data types in scope for INSDC, calls for a much higher breadth of technical maturity than associate

membership, which is focused on small, specific scientific domains of interest with limited volumes of available data.

Data operations examines the maturity of an institution to handle data at the scale and throughput of INSDC. Specifically, institutions need to demonstrate maturity in *secure management of data* (i.e., private, pre-released, or explicitly withdrawn data are protected from intentional or accidental public release) and in *incoming and outgoing data and metadata flow*. Data flow deals with an institution's ability to assign accessions, exchange data with other INSDC partners at their chosen level of scale (it is anticipated that Associate Members will exchange smaller volumes of data than full members), and confirm that data it holds are correctly synchronized with the rest of INSDC.

Communications and engagement gauges the ability of an institution to act as a representative of INSDC internally with other INSDC Members (*technical engagement*), as well as externally on the global scientific stage (*community outreach, representation and support*). To demonstrate maturity in this area, institutions must show they are ready and able to participate fully in INSDC discussions (e.g., INSDC technical meetings) and have a track record of representing the work of their resources to the international community – for example through presentations and outreach to potential users. The institution should also be able to demonstrate an active engagement with users and stakeholders through support or helpdesk processes, as well as an ability to respond to user feedback.

Quality assurance looks at the ability of potential INSDC Members to ensure the *data quality* of their resources meets INSDC standards. This includes checks on the syntactic validity of the data (e.g., file format checks, such as verifying FASTQ files are indeed readable as FASTQ files by common bioinformatics tools) and scientific quality checks (e.g., minimum metadata criteria or annotation correctness assessments). Self-assessment in this area also includes an evaluation of data maintenance mechanisms, such as those that check data statuses. For example, if INSDC endorsed vocabularies change (e.g., a new taxonomy categorisation is used), INSDC databases commit to updating records accordingly to ensure that the data available from INSDC partners is adherent to the latest INSDC-endorsed standards.

Evaluation

The INSDC Maturity Model uses a series of objective criteria – or indicators – for all of the categories in each area that are arranged in increasing levels of maturity. Associate Members are expected to demonstrate a level of maturity aligned to the scale and scope of the data they support, while those seeking full membership will be assessed against their ability to show these activities to the full breadth of INSDC data. Demonstrating that an institution has sufficient capabilities in place and that these capabilities are sufficiently mature will be a prerequisite for INSDC membership at either level.

The first version of the INSDC Maturity Model is available upon request from https://www.insdc.org/contact/. This version is intended to support initial membership evaluations, and whilst we do not expect the capabilities this model assesses to change, the way in which indicators are arranged, specific details of the maturity levels, and the language used to describe the indicators may change significantly in revisions of this model in efforts to improve clarity and support greater ease of use by prospective members. We welcome all feedback on the model from any institution using it to evaluate the maturity of a sequence database resource.

Appendix

Membership Levels

INSDC Members support the deposition, discovery, retrieval, analysis, and reuse of nucleotide sequence data. INSDC Members will be named on the INSDC website. There are two levels of INSDC membership: Full and Associate.

1. Full Members collect, archive, and present for search, analysis, and retrieval all their nucleotide sequence data types in scope for INSDC as defined by the INSDC Executive Committee (into which they provide representatives) and documented on the INSDC website including providing access to all data submitted to other Members to enable reuse of the complete INSDC dataset.

2. Associate Members collect, archive, and present a subset of sequence data for search, analysis and retrieval that is limited in scope either by region, scientific domain, and/or by data type. Associate Members provide access to relevant INSDC datasets consistent with the scope of their collections.

Additional Membership Expectations

In addition to demonstrating the mature capabilities that are required, organizations intending to apply for membership in INSDC are expected to have worked with an existing INSDC Full Member to establish bilateral data sharing arrangements, such as data submissions brokering, outside formal INSDC membership. Doing so allows prospective Members to gain experience with capturing, processing, and sharing sequence data on a large scale, and is an effective way to gain practical experience with capabilities that are aligned to the INSDC Maturity Model. Demonstrating maturity in these areas is critical for success as a Member of INSDC.