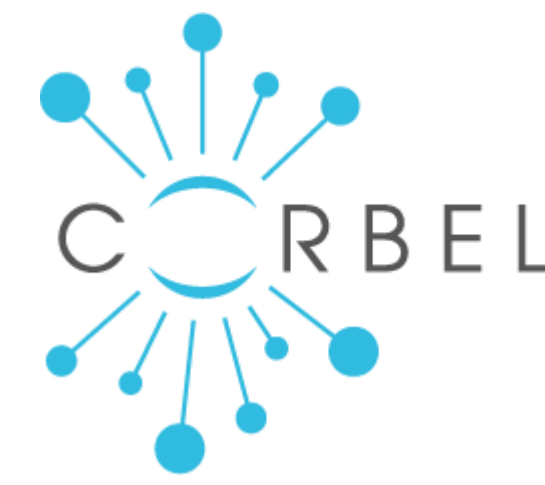# EOSC-LIFE

WP6 T6.3 COMMON PROVENANCE MODEL

FOR PROCESSING BIOLOGICAL MATERIAL, DATA AND WORKFLOWS

The goal of the task deliverable is to create a provenance information model and its usage requirements for the biotechnology domain covering the whole process chain, from the source of biological material, through its processing, analysis, and all steps of data generation and processing.

FEATURES OF THE MODEL:

1. Enabling **effective assessment** of **quality** and **fitness for purpose** of the objects provided, such as biological material and data;
2. Supporting **reproducible research** by exacting the capture of all relevant information;
3. Tracking **error propagation** within scientific results;
4. Tracking the **source of biological material** in order to prevent fabrication of data and enabling the notification of subjects in case of relevant incidental findings;
5. Propagating **withdrawal** of or **changes** to an informed consent along the process chain;

The provenance model is subject of standardization at the International Organization for Standardization.

Standard coverage.

Sample Acquisition, Processing, Transport, and Storage Provenance (IS) (Part 3)	Data Generation Provenance (NGS, mass spec, ...) (IS) (Part 4)	Data Storage and Processing Provenance (CWL, ...) (IS) (Part 5)

Provenance Information Management Requirements (TS) (Part 1)

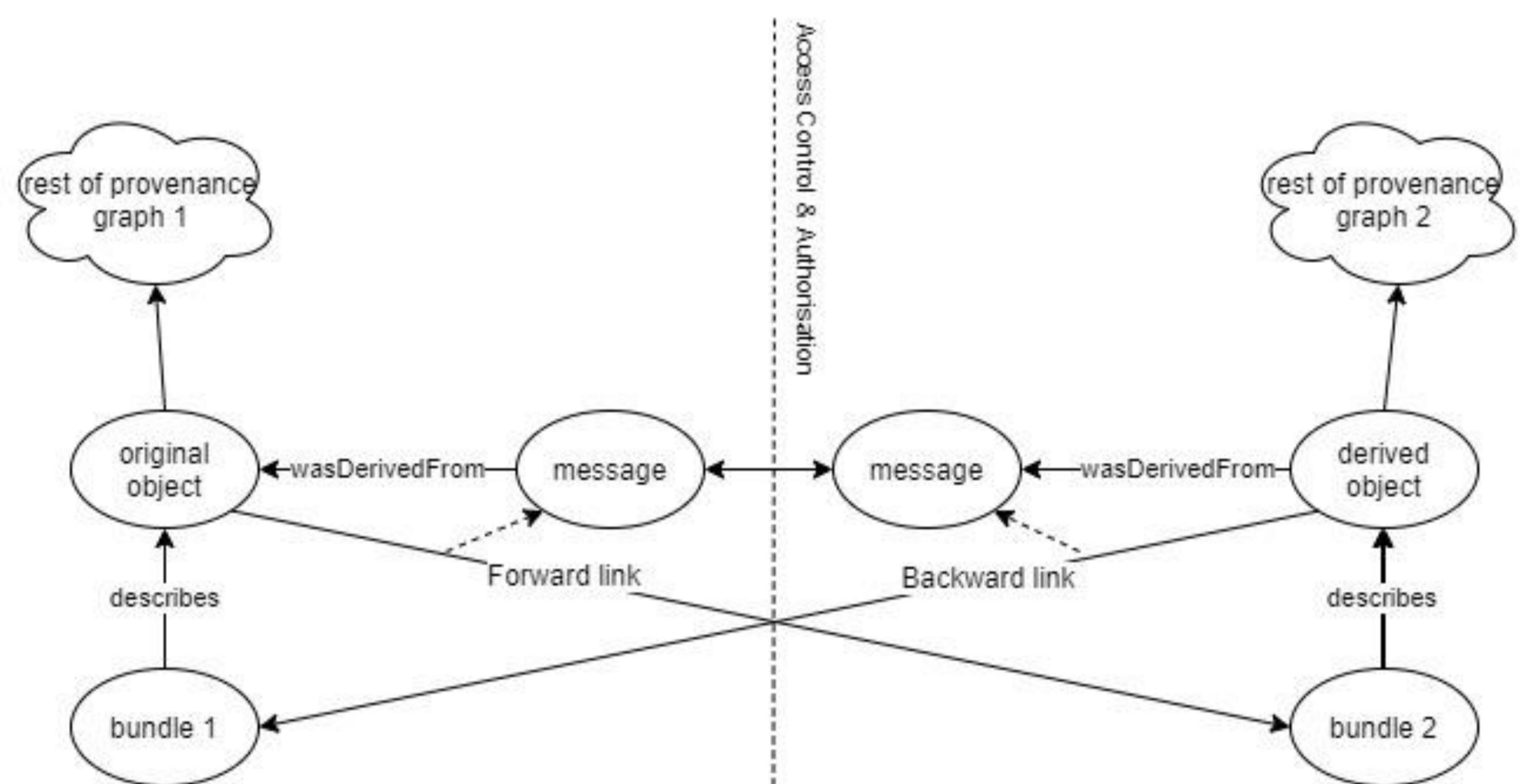| Common Provenance Model (TS) (Part 2) |
|---|

Security Extensions (TS) (Part 6)

Proposed structure of the standard TC276/WG5 23494

TASK LEADERS, CONTRIBUTORS AND INTERACTIONS

Petr Holub, Isabelle Perseil, Jörg Geiger, Rudolf Wittner, Carole Goble, Heimo Müller, Stian Soiland-Reyes, Francesca Frexia, Cecilia Mascia, Elliot Fairweather, Luca Pirredu, Gianluigi Zanetti, Hiroki Nakae, Caterina Strambio, Josh Moore, David Grunwald, Jason Swedlow
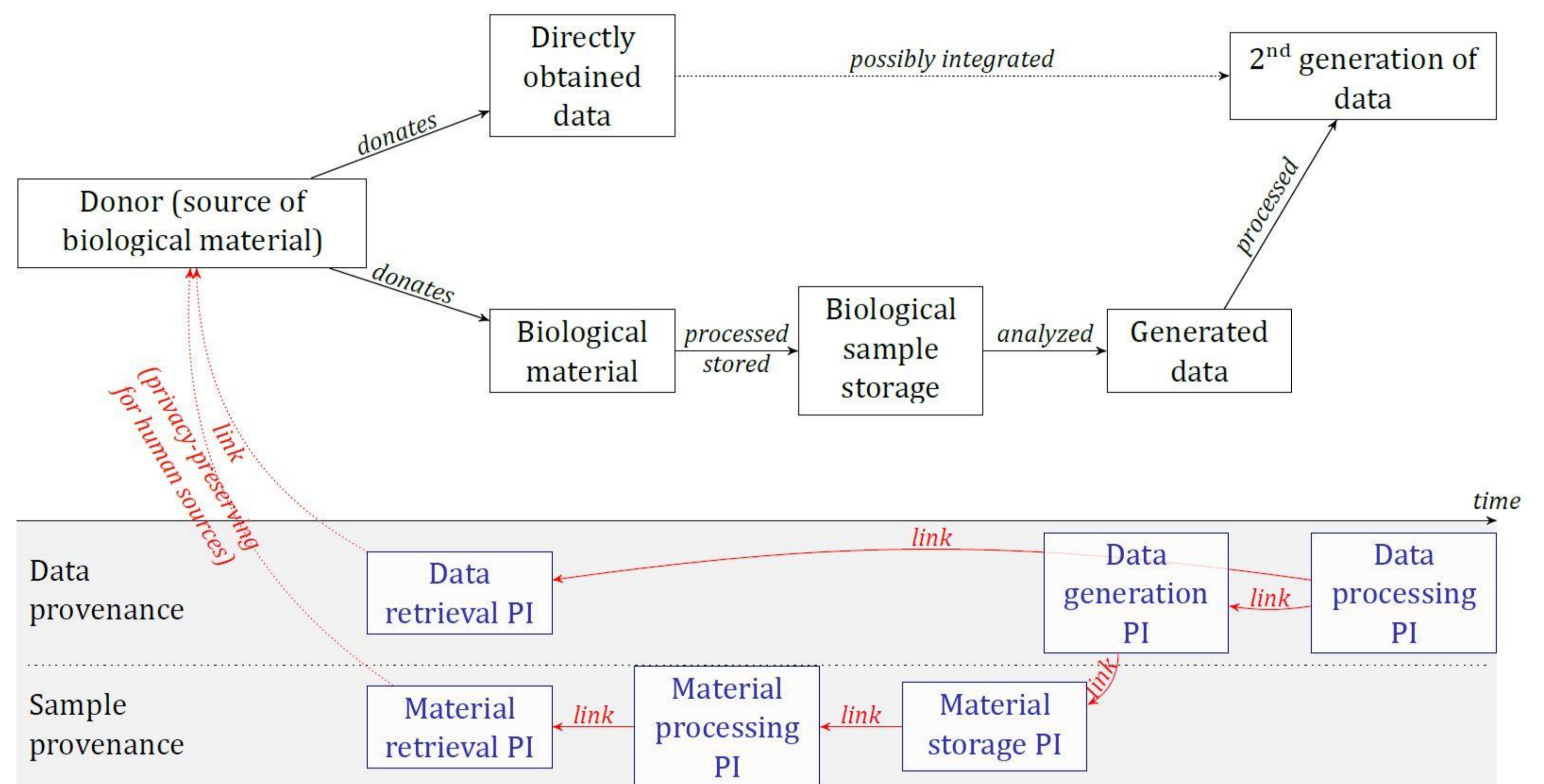
General schema of a distributed provenance model

FOCUS AREAS:

1. Applying **W3C PROV to describe** all phases of biomedical research and its enrichment by **new types of structures** (e.g. relations, entities, …) to capture common objects.
2. Definition of provenance templates as **common representation** of typical scenarios.
3. Interconnecting **distributed provenance** to enable processing of provenance stored within multiple organisations with support for **opaque provenance components** .
4. Full **syntactic and semantic interoperabili**ty of captured provenance.
5. Rigorous **formal verification process** of provenance instance validity (provable compliance with the model).
6. Access control, integrity and non-repudiation, protection of privacy.