Manifesto of the Personal Health Train consortium for a concerted action to dovetail FAIR data and services in healthcare and biomedical science

The undersigned represent major public and private stakeholders in healthcare and biomedical research in the Netherlands, who are currently developing the ‘Personal Health Train’ (PHT). This is a distributed infrastructure that enables the use and reuse of health data for the benefit of individuals and society. The PHT aims to advance healthcare and biomedical science through a shared infrastructure to support data management (healthcare data, health & lifestyle data, and biomedical research data), data analysis, and medical decision making. The PHT constitutes a crucial element of various national initiatives, including Health-RI, DTL, and ELIXIR-NL, and it will act as a GO FAIR Implementation Network. This manifesto is a formal statement of the principles that guide the PHT development.

Empowering citizens

We recognise the increasingly important role of citizens in the prevention and treatment of disease. The PHT empowers citizens to control access to their personal health data, to provide dynamic consent to the use of their data in healthcare and scientific research, to partner with healthcare providers in medical decision making, and to support patient-initiated research. It respects the autonomy and privacy of citizens and implements important aspects of responsible data science.

We therefore commit to the following guiding principles for the development of the PHT:

- **Control over data.** The PHT empowers citizens, public or private organisations to manage, safeguard, and share their data for use in healthcare or scientific research.
- **Reusable personal health data.** The PHT is a shared digital infrastructure based on standards and protocols adhering to the FAIR principles (i.e., digital resources are Findable, Accessible, Interoperable, and Reusable).
- **Distributed and federated solutions.** The PHT architecture relies on distributed and federated learning and decision support where possible. Data stay where they are and are processed at their location of origin, unless distributed solutions are not (yet) available or do not suffice.
- **Responsible use of personal health data.** We will act to enable and ensure the responsible use of personal data by adopting international principles and regulations, including the FACT principles (Fairness, Accuracy, Confidentiality, Transparency), privacy-by-design, privacy-by-default, and the General Data Protection Regulation (GDPR).
- **Ethics-by-design.** We commit to optimise the facilities to judge the ethical aspects of research questions and enable blocking and reporting of studies that abuse personal data.
- **An open ecosystem for innovation in health and well-being.** Everyone subscribing these guiding principles can contribute to the development of the PHT. We will adopt and develop open standards and protocols. We strive to avoid single-vendor solutions that create single points of failure for any critical component of the shared infrastructure. The core infrastructure will be a common and public good.
- **Registration at the source.** Data should be captured only once, promoting efficiency by avoiding repetitive work. Copying or moving source data by authorised individuals should be made explicit in the data provenance and limited as much as possible.
- **Machine-readability at the core.** We focus on creating machine-readable and interpretable data, metadata, workflows, and services, aiming for maximal interoperability between diverse systems including electronic patient records. Machine-readable data will be accompanied by human-readable versions in different languages for different audiences (professionals, citizens).

Though we appreciate diversity, especially in the research field, we consider this joint statement a way to speak with one voice on a number of critical issues that are of generic importance and on which we feel we have reached consensus. We will therefore coordinate our investments in and support of the technological and social developments in the distributed management and analysis of personal health data.