THE GFBIO DATA SUBMISSION AND BROKERAGE SYSTEM

Ivaylo Kostadinov, Marc Weber, Deniss Marinuks, Jimena Linares, Frank Oliver Glöckner
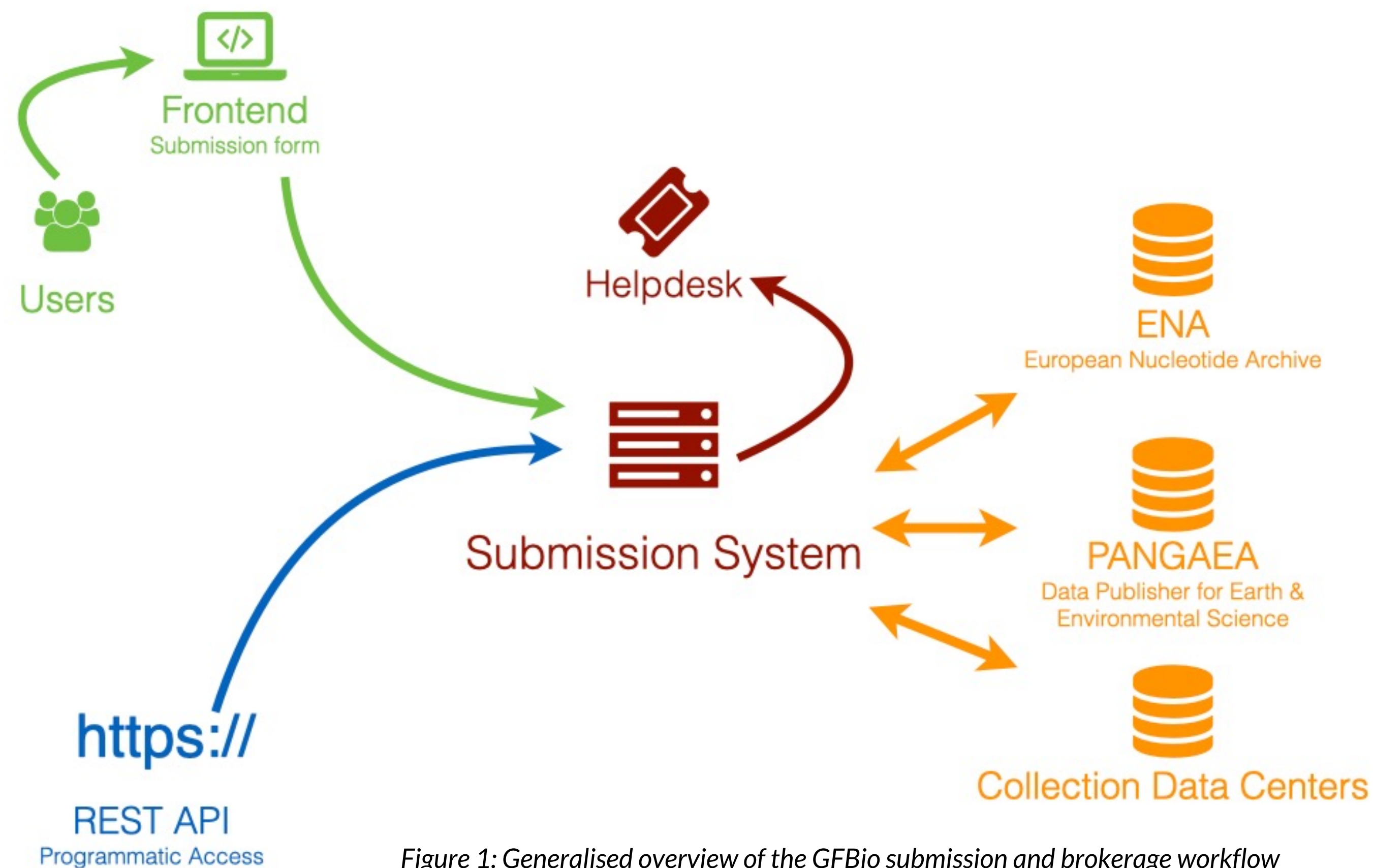
Figure 1: Generalised overview of the GFBio submission and brokerage workflow

THE CORE SERVICE

Since late 2016 GFBio operates a data submission service, which allows users to submit data to any of the originally nine, now ten, GFBio data centres for long-term preservation and publication. The service consists of a web-based frontend and a modular backend, which provides features like validation, automated transfer to some of the data centres and embargo reminders. The service is tightly integrated with the GFBio Helpdesk – all communication with the submitter, including automated messages and comments by curators at the GFBio data centres, is documented in dedicated submission tickets.

https://submissions.gfbio.org

https://www.gfbio.org/data-centers

KEY FEATURES

IN NUMBERS

- Over 230 data submission requests
- Over 100 unique submitters
- Over 40 returning submitters
- Over 4TB of research data brokered

- Expert support for metadata standardization (e.g. ABCD, DwC, MIxS, ENVO)
- Manual (GUI) and programmatic (API) operation
- Complex data sets are split and deposited in the most appropriate data centres, then interlinked via persistent identifiers, e.g. DOIs (Fig. 2)
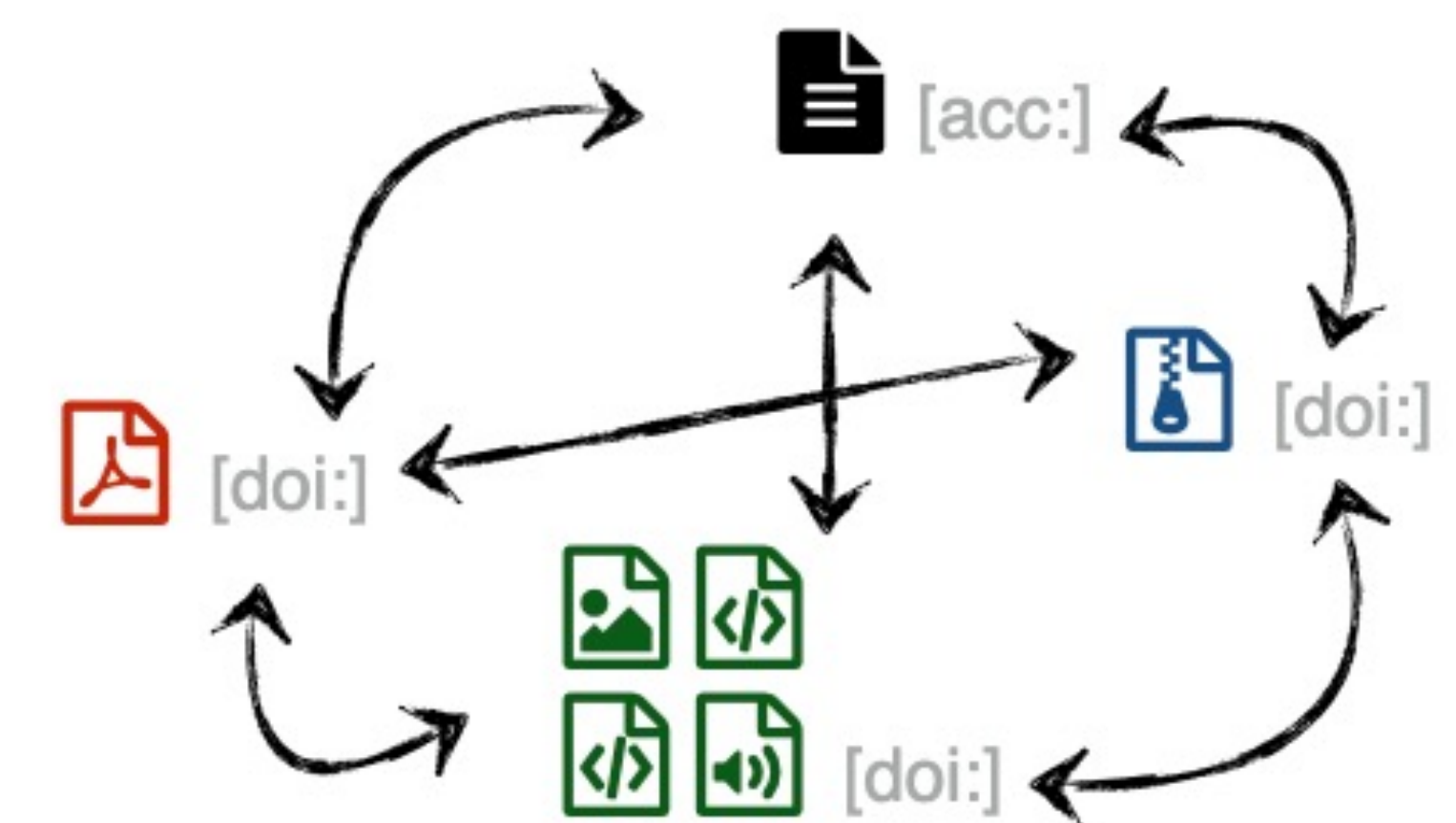- High level of automation and resilience

Figure 2: A stylised depiction of a network of research data output. Different data types, belonging to the same data set are deposited in dedicated data centres and interlinked via persistent identifiers.