

## Integrating Activities for Advanced Communities



### D3.3 Virtual Access Assessment Report 1

Project No.871120– INTERACT

H2020-INFRAIA-2019-1

Start date of project: 2020/01/01

Duration: 48 months

Due date of deliverable: 2021/12/31 (M24)

Actual Submission date: 2022/01/17

Lead partner for deliverable: UOULU

Author: Hannele Savela

Dissemination Level		
<b>PU</b>	Public	X
<b>PP</b>	Restricted to other programme participants (including the Commission Services)	
<b>RE</b>	Restricted to a group specified by the Consortium (including the Commission Services)	
<b>CO</b>	Confidential, only for members of the Consortium (including the Commission Services)	

---

## Table of Contents

<b>Publishable Executive Summary.....</b>	<b>3</b>
<b>1. Virtual Access Assessment Report.....</b>	<b>4</b>
1.1. Background.....	4
1.2. Virtual Access Board's assessment results.....	5
1.3. Conclusions and next steps.....	5
<b>Appendix 1 .....</b>	<b>6</b>

---

## Publishable Executive Summary

Virtual Access (VA) means free on-line access for users to metadata and datasets collected and maintained by the organizations and research infrastructures offering VA in INTERACT III (GA No. 871120). In total, VA will be offered by 29 INTERACT partner organizations by the end of the project period in 2023. At the moment, ten organizations have already joined VA provision through the new [INTERACT Data Portal](#) that was launched in March 2021. The other nineteen VA providers will join the data portal during the second half of the project period, and until then offer VA through the [previous version of the Virtual Access Single Entry Point](#) that was established in INTERACT II.

Virtual Access offered during the first half of the project period was assessed by the Virtual Access Board (VAB), which recommends e.g. that the Virtual Access activity in INTERACT continues and put special focus on enabling machine interaction, focussing on free and open data, emphasises the need for unique identifiers and for clear citation information on datasets. In addition, VAB recommended INTERACT to consider a formal FAIRness assessment of the data access infrastructure to discover additional tasks towards data FAIRness. The INTERACT Guidebook that provides both high level introduction for readers not familiar with the concepts of data management as well as more in depth description of the technical interfaces needed was assessed as very appropriate and useful by VAB.

The recommendations suggested by the VAB are considered essential to adhere to the FAIR guiding principles for discovery metadata. The recommendations of the VA Board will be taken into account and implemented in the second half of the project period, followed by the second VA Assessment in 2023.

---

## 1. Virtual Access Assessment Report

### 1.1. Background

Virtual Access (VA) means free on-line access for users to metadata and datasets collected and maintained by the organizations and research infrastructures offering VA in INTERACT III (GA No. 871120). In total, VA will be offered by 29 INTERACT partner organizations by the end of the project period in 2023. At the moment, ten organizations have already joined the VA provision through the new [INTERACT Data Portal](#) that was launched in March 2021. The other 19 VA providers will join the data portal during the second half of the project period, and until then offer VA through the [previous version of the Virtual Access Single Entry Point](#) that was established in INTERACT II.

The INTERACT Data Portal is based on metadata harvesting from organizations and sites with end points to real data on various topics on e.g. earth sciences and ecosystems. The number of datasets presented in the portal currently includes over 1500 datasets, and the number of datasets is growing day by day. The data presented includes both near real-time observation datasets and unique historical, retrospective data digitized and provided for open access.

In addition to the development of the new INTERACT Data Portal for Virtual Access provision, several guiding documents have been developed to help the VA providers to join with the VA provision through the data portal. These guiding documents include the INTERACT Metadata Guidelines, and the INTERACT Guidebook on how to make metadata available in the INTERACT Data Portal. These guiding documents are publicly available at <https://eu-interact.org/wp-3-giving-access-to-the-arctic/>

In addition to the guiding documents, there are several structures in place in the INTERACT project organization to ensure collaboration and communication across work packages and activities related to data provision and management, and to safeguard the successful development of the VA provision in INTERACT. These structures include the INTERACT Data Team and Data Safeguard.

The Virtual Access was assessed by the Virtual Access Board (VAB) based on the following materials and background information:

- INTERACT Data Portal
- INTERACT VA Single-Entry Point website (from INTERACT II)
- WP11 (VA) Technical Report for the 1<sup>st</sup> Reporting Period (1.1.2020-30.9.2021)
- INTERACT TA/RA/VA Quality Assurance Programme
- INTERACT Metadata Guidelines
- INTERACT Guidebook: how to make metadata available in the INTERACT Data Portal

The composition of VAB was set by the INTERACT Data Safeguard (WP1), and it comprises of persons with extensive experience on data management and deep insight on the opportunities and challenges related to providing free online access to data.

The members of the VAB are:

- Øystein Godøy, Norwegian Meteorological Institute, Oslo (chair, INTERACT Data Safeguard)

- Kirsten Elger, GFZ German Research Centre for Geosciences, Potsdam
- Markus Fiebig, Norwegian Institute for Air Research, Kjeller
- Peter Pulsifer, Carleton University, Ottawa

## **1.2. Virtual Access Board's assessment results**

The INTERACT Virtual Access Board (VAB) has evaluated the materials provided, and recommends in their assessment (Appendix 1) that the Virtual Access activity in INTERACT:

1. continues the current activity of enabling machine interaction with the data catalogues of INTERACT stations with the ambition of moving stations from the old virtual access portal to the new,
2. puts in place mechanisms for quality control and ongoing maintenance of metadata records, e.g. for dead links for data access,
3. continues to emphasise the principles of free and open data, including the [IASC principles of ethically open data](#),
4. consolidates the technology stack and guidance documentation already developed for exchange of discovery metadata,
5. emphasises the need for unique identifiers (preferably persistent) on all metadata records and datasets exposed,
6. promotes the use of widely used persistent identifier systems, especially DOI, ORCID, ROR
7. emphasises the need for clearly stated licenses on datasets, ideally included in the machine-readable metadata
8. promotes and strengthens the requirement for application of controlled vocabularies for specific elements of the discovery metadata model used,
9. emphasises the need for clear citation information on datasets.
10. starts an exercise towards specifying vocabularies to be used in metadata exchange, in order to work towards the “rich metadata” criterion on data FAIRness.
11. considers a formal FAIRness assessment of the data access infrastructure in order to discover additional tasks towards data FAIRness.

These recommendations are considered essential to adhere to the FAIR guiding principles for discovery metadata.

## **1.3. Conclusions and next steps**

The feedback and recommendations provided by the VA Board in their assessment on INTERACT Virtual Access provision will be taken into account and implemented during the second half of the project period, leading to the second VA Assessment in the end of the INTERACT III in 2023.

---

## Appendix 1

# INTERACT III

## Virtual Access Assessment #1

The INTERACT Virtual Access Board (VAB) has 4 members and has evaluated the material provided. Members of the VAB are Kirsten Elger (GFZ German Research Centre for Geosciences, Potsdam), Markus Fiebig (Norwegian Institute for Air Research), Peter Pulsifer (Carleton University, Ottawa) and Øystein Godøy (Norwegian Meteorological Institute). This statement includes the first Virtual Access Assessment for INTERACT III. The statement is based on the input provided through the periodic technical reports for reporting period #1 for WP 3 and 11 as well as the INTERACT Guidebook on Making metadata available for virtual access in INTERACT Data Portal and Deliverable D3.2 “How to get the data INTERACT metadata tag pilot technical manual”.

The INTERACT Virtual Access Single Entry Point is now changed to <https://dataportal.eu-interact.org/> contrary to earlier when this was a web page in the project website at <https://eu-interact.org/accessing-the-arctic/virtual-access/>. The latter provides a good overview of the current status of virtual access within INTERACT, but the most important element is the reference to the first URL, i.e. the actual implementation of a single entry point based on interoperability standards. This new entry point is now named the INTERACT Data Portal and contains more than 1500 datasets from 10 INTERACT stations. Datasets include both real time data flows and historical datasets, sometimes published with DOI. When fully functional, information on datasets from 29 stations should be available in the system. According to the periodic report of WP 3, 14 stations are still providing access to data through the old virtual access portal. Among these 14, 7 are in the final stages of preparing end points for metadata harvesting. This is reflected by many stations not having any data visible in the new portal yet.

To support stations, guidance material has been developed. This includes as mentioned above a brief deliverable D3.2 how-to manual and a more comprehensive INTERACT Guidebook. This material provides both a high level introduction to concepts for readers not familiar with the concepts of data management as well as more in depth description of the technical interfaces needed. Given the diversity of the INTERACT community, the guidance material developed seems very appropriate and useful, although it is assumed that the data team has identified a number of issues to improve in future versions as part of the dialogue with the community. This is normal in such circumstances.

INTERACT has chosen to base the Data Portal on CKAN software and relying on the schema.org vocabularies and protocols adhering to the principles outlined by Earth Science Information Partners (ESIP) sub-group called *Science on schema.org* and the DCAT-AP for metadata harvesting. This looks like a good and sound approach at this point in the development and has increasingly broad adoption within the polar data community. It should however be noted that the approach should be considered a bare minimum solution concerning the richness of metadata and also in the perspective of FAIRness as it basically only addresses the issue of discovery metadata and does not address FAIRness at the data level. The FAIR model explicitly recognized this nested approach by allowing discovery metadata to link to more detailed metadata and actual data collections (see principle I3). However, again from the perspective of the previous state of metadata exchange within the INTERACT community this is a huge step forward. As such,

---

the INTERACT Virtual Access approach has made its first steps into the FAIR space, addressing F (Findable) and A (accessible).

In this context and based on examination of the data portal, VAB would like to recommend that the virtual access activity in INTERACT:

1. continues the current activity of enabling machine interaction with the data catalogues of INTERACT stations with the ambition of moving stations from the old virtual access portal to the new,
2. puts in place mechanisms for quality control and ongoing maintenance of metadata records, e.g. for dead links for data access,
3. continues to emphasise the principles of free and open data, including the [IASC principles of ethically open data](#),
4. consolidates the technology stack and guidance documentation already developed for exchange of discovery metadata,
5. emphasises the need for unique identifiers (preferably persistent) on all metadata records and datasets exposed,
6. promotes the use of widely used persistent identifier systems, especially DOI, ORCID, ROR
7. emphasises the need for clearly stated licenses on datasets, ideally included in the machine-readable metadata
8. promotes and strengthens the requirement for application of controlled vocabularies for specific elements of the discovery metadata model used,
9. emphasises the need for clear citation information on datasets.
10. starts an exercise towards specifying vocabularies to be used in metadata exchange, in order to work towards the “rich metadata” criterion on data FAIRness.
11. considers a formal FAIRness assessment of the data access infrastructure in order to discover additional tasks towards data FAIRness.

These recommendations are considered essential to adhere to the FAIR guiding principles for discovery metadata.