**Data Management Plan (DMP):** Figure 2 summarizes the proposed flow of data within the CPH overall, whereas Figure 3 demonstrates the general data capture and digital preservation at the smaller, outlying institutions. After cleaning (Fig. 2), data will be exported from the CPH database to the PBIN (Pacific Basin Information Node of the United States Geological Survey), which is a component of the Pacific Biodiversity Information Forum (USGS), and a data provider to GBIF (Global Biodiversity Information Facility). Bishop Museum (including BISH) has long provided its database information directly to PBIN/PBIF, whose mission includes making scientifically sound data from the Pacific Basin accessible online.

Figure 2. Proposed general data management system for the Consortium of Pacific Herbaria.
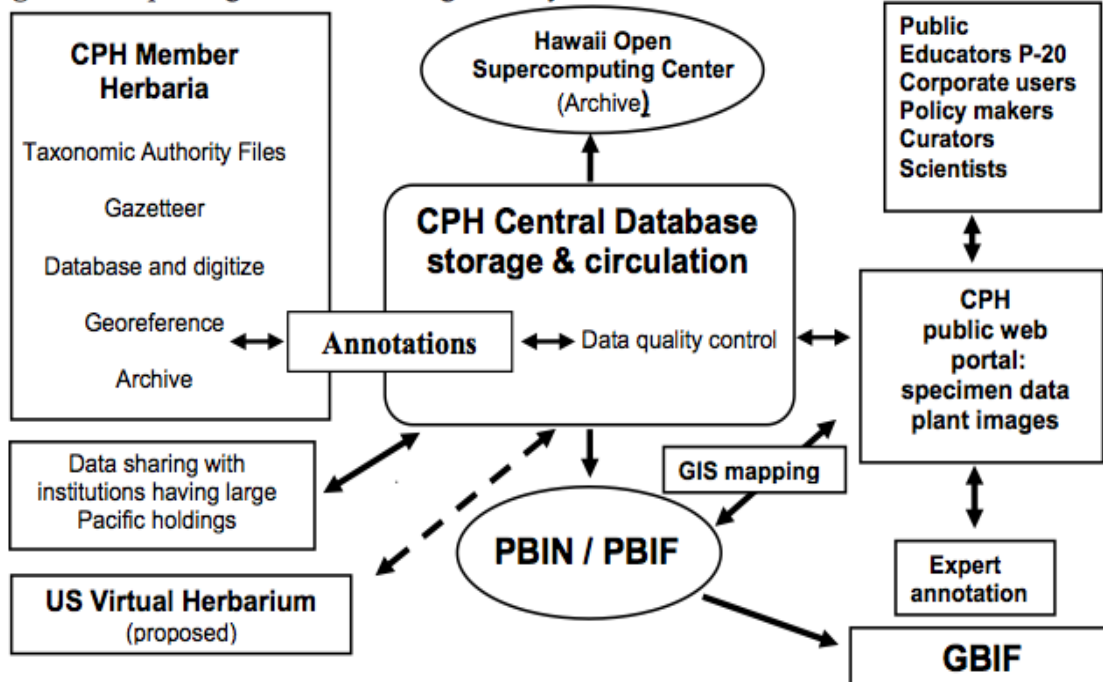
James (co-PI), and Thomas and Pyle (senior personnel), have extensive experience in the design, implementation, and upgrading of database systems. For the Taxonomic Authority File the familial taxonomy generally will follow Smith et al. (2006) for ferns, Mabberley (2008) for gymnosperms and most generic concepts, and APGIII (2010) for angiosperms but with the modifications of Nickrent et al. (2010). Database design will follow well-established protocols such as the Darwin Core and TAPIR metadata standards such as DiGIR (BIS 2010), which are used at BISH, PTBG, and HAW. Geographical names, of which there are many variants in the Pacific, will be standardized using Motteler (2006).

Each herbarium will enter its specimen data, which will be forwarded monthly to the CPH Central Database (Figs. 2-3). After cleaning for errors or inconsistencies, incoming data will be forwarded to PBIN/PBIF, which provides data to GBIF, and be released to the public via the CPH

11

public portal (Fig. 2). An important feature of the data management will be a bidirectional information flow of specimen annotations (Figs 2-3). This requires CPH members to forward all specimen annotations to the Central Database, in addition to updating their own in-house database. Although the "filtered push" of duplicated data across institutions (Wang et al. 2009) is nearing proof of concept stage, it is not easily implemented at this time. We will stay aware of its development and consider its implementation, if feasible, at a future date.

The smaller, outlying herbaria house many unique collections that mostly are not digitally imaged or databased. Institutions such as SUVA and GUAM have large numbers of images to capture. Figure 3 summarizes the capture and digital preservation of data for these institutions. Barcoding may be included in the smaller institutions as part of work flow depending on their protocols. Thomas (senior personnel) will travel to each institution during the first year to set up the database (including Taxonomic Authority File) and portable digital imaging stations, and provide training of personnel on site in the use of the database and imaging stations. While on site he will also carry out a conservation review for each herbarium, checking for concerns related to insect pest management, proper storage of specimens (air conditioning, sealed cabinets, non-leaking jars for spirit material) and give recommendations on specimens in need of conservation (re-mounting, splitting into two or more sheets, etc.). The Waimea Arboretum on Oahu will be used as a test case to set up the database and digital imaging stations to be used in other herbaria.

Bandwidth capacity recently increased substantially for the University of Fiji (155 mbps), and a high bandwidth fiber optic cable will soon be implemented for Guam and the Northern Mariana Islands using the Asia-American Gateway. Collections data from the smaller herbaria can be sent electronically without concerns. However, given the large size of image files, data will be sent on a 1 terabyte hard drive monthly (Fig. 3), the approach used by the JSTOR Plants Initiative for its capture of digital images of type specimens. Each smaller herbarium will have 2 hard drives; one on site, and one in transit. Herbaria with less than 1,000 specimens will send the data on the hard drive only once. HAW and PTBG have adequate storage capacity to handle the respective number of digital images. BISH has virtually exhausted its data storage capacity and will require new rack networks (see budget justification). The CPH database will be backed up nightly on the CPH Central Database and archived weekly to the Maui High Performance Computing Center (Fig. 2), whereas BISH currently archives its data nightly to disk and weekly on tape.

Figure 3. Proposed data capture and preservation at smaller CPH institutions. Label data can be databased either from the physical specimen or from the digital image for efficiency.
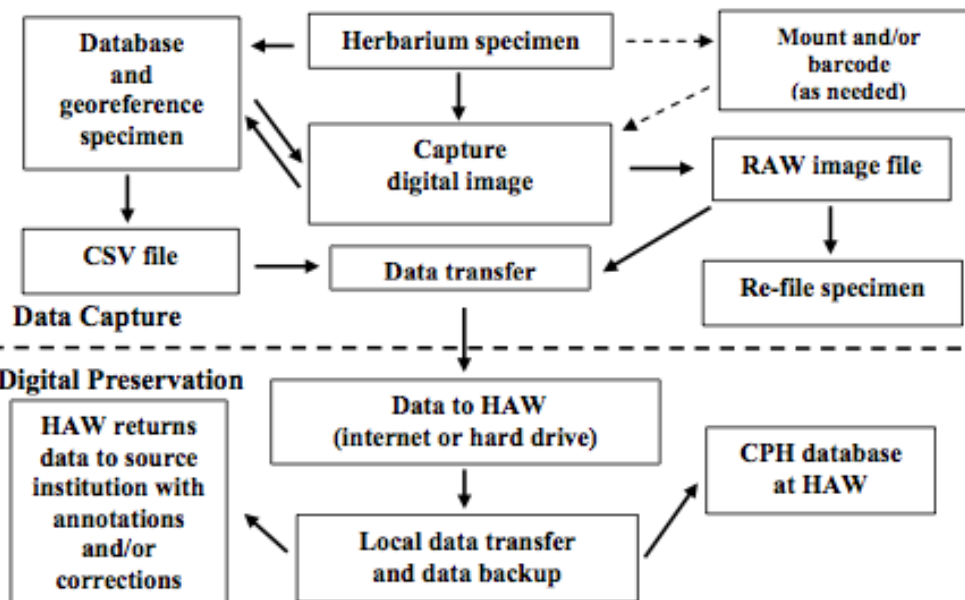
CPH database development: James, Thomas and Snow will oversee development of the CPH database and web portal overall. We request funding for a 0.5 FTE for a graduate assistant and as an on-contract web designer at HAW for the continued development of the CPH database and web portal. Three skill sets are needed: programming, web interface design, and graphics design.