# PIM and Personality: What do our Personal File Systems Say About Us?

**Charlotte Massey, Sean TenBrook, Chaconne Tatum, Steve Whittaker**
University of California at Santa Cruz,
High St, Santa Cruz, 95064.
{cmassey, stenbroo, ctatumdi, swhittak}@ucsc.edu

## ABSTRACT

Individual differences are prevalent in personal information management (PIM). There is large variation between individuals in how they structure and retrieve information from personal archives. These differences make it hard to develop general PIM tools. However we know little about the *origins* of these differences. We present two studies evaluating whether differences arise from *personality traits,* by exploring whether different personalities structure personal archives differently. The first exploratory study asks participants to identify PIM cues that signal personality traits. While the aim was to identify cues, these cues also proved surprisingly accurate indicators of personality. In a second study, to evaluate these cues, we directly measure relations between structure and traits. We demonstrate that Conscientiousness predicts file organization, particularly PC users' desktops. Neurotic people may also keep more desktop files. One implication is that systems might be customized for different personalities. We also advance personality theory, showing that personal digital artifacts signal personality.

## Author Keywords

Personal Information Management; file systems; individual differences; personality.

## ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

## INTRODUCTION

Personal Information Management (PIM) is a fact of life; all of us spend time every day organizing personal information, both to facilitate its later retrieval and to manage our tasks. One repeated observation in PIM is *individual differences* in how people structure and retrieve personal information. However less attention has been paid to the *reasons* for these differences. This is an important question because differences in organizational structure have direct consequences for retrieval [5], task management [36], and for stress [8]. And despite improvements in desktop search, search has not replaced the need to actively organize personal files. Navigation through one's personally organized file system is still people's preferred way to find their information, underscoring the importance of understanding different organizational strategies [4].

There are clearly many potential causes of individual differences in PIM. One classic individual difference is between *filers* and *pilers*. Prior work speculates that these differences partially arise from different work responsibilities [3, 36] or individual cognitive processing style [19]. However, so far no PIM research has directly explored whether differences arise from *personality traits*. In personality theory, traits are defined as habitual patterns of behavior, thought, and emotion [26]. The consensus in personality theory is there are 5 main personality traits of Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism [15, 21]. Such traits may directly influence how people organize information. We can therefore ask whether Neurotics organize personal information differently, and whether there are effects of other personality factors such as Conscientiousness or Openness.

A new approach to this question is suggested by recent work on personality examining relationships between people's personal *physical* environments and their personality. This approach explores whether personal environments, e.g. bedrooms or offices, directly signal their owners' personality traits. [16, 17] shows that strangers inspecting an unoccupied college dorm room can reliably determine how Open and Conscientious the owner is, based on *cues* such as travel souvenirs or a wide variety of books and music. Inspectors are moderate at judging other traits, apart from Agreeableness. However there are sometimes discrepancies between what room inspectors *think* are accurate trait cues, and what *actually* are reliable cues. So for example, while inspectors *thought* that towels strewn on bedroom floors indicated the owner's Disagreeableness, *in reality* formality of décor is a better cue.

Although [17] shows reliable relations between personal physical environments and certain personality traits, less research has explored the *digital* domain. We know that public aspects of information spaces, such as Facebook profiles or personal web pages are deliberately designed to

convey identity [1, 11, 18, 29, 35]. However does personality also affect how we organize less public aspects of our personal digital collections that others aren't intended to see? Do our personalities affect how we organize personal files?

We examine whether *organization of personal archives* reflect personality. We ask, for example, whether Conscientiousness people have smaller overall archives, and do they more actively use folders and the desktop? Do Open individuals have large amounts of pictures and music and so forth? We focus on personal archives including documents, music and folders. We exclude email organization as this largely involves processing information sent by others, rather than generating, finding and organizing personally relevant information [37]. We address the following questions:

- *Is there a relationship between people's personality and their organization of personal digital archives?*

- *How does personality affect organizational strategies?*

- *Can strangers identify structural cues that signal personality based on how a person organizes their personal archives?*

We use two different methods. Our first exploratory study presents people ('analysts') with structural information about how a stranger's personal digital archive is organized. We ask these analysts to first identify what structural aspects of the archive provide *cues* to the archive owner's personality, e.g. an organized Desktop reveals Conscientiousness. We then ask analysts to *predict* the owner's personality traits using these cues. In the second study, we test the validity of the structural cues identified in the first study. Participants complete a standard personality survey. We analyze the structure of their personal digital archive, using a program that generates structural information including number of files, files/folder, use of the desktop etc. We then assess whether and how owners' personality predicts objective structural properties of their personal archives.

Overall we demonstrate relations between PIM and personality. The first study generated a set of cues about which aspects of PIM structure signal personality, surprisingly showing that people may be able to use such structural information to reliably predict overall personality traits. The second study found that Conscientiousness and Neuroticism predict objective characteristics of people's personal archives, in particular desktop usage.

## RELATED WORK

### PIM and Individual Differences
PIM is an active research area, with new theoretical frameworks [23, 37] building on pioneering empirical studies examining how people organize their information [5, 25]. More recent work demonstrates that organizational structure predicts how successful people are at refinding

personal information [5]. The importance of PIM for daily work means that many new software systems have been developed to improve organization and retrieval of personal information. These include sophisticated search tools [10] and faceted classification [7]. However these new organizational tools have not replaced the hierarchical file system: users still prefer to organize and retrieve information by navigating their personal file hierarchy [4, 34].

Malone [25] documented individual differences in how people organize paper archives that he called *filing* and *piling*. *Filers* organize collections into fine-grained subclassifications of semantically related items, whereas *pilers* make shallower collections involving large classes of heterogeneous objects related by temporal rather than semantic association. Email research has found similar strategies. Some people create complex folder structures to organize incoming messages (filers), while others leave many messages in the inbox relying on inbox scanning for retrieval (pilers) [36]. Although these simple strategy classifications have subsequently been refined [12, 19], there is consensus that there are distinct individual differences in organizational strategies. Work exploring the relations between organizational strategy and retrieval has shown that shallower broader organizations promote more effective and successful retrieval [5]. However there has been little systematic examination of the causes of these individual differences, which we explore here.

### Personality Traits and Their Relation to Physical Environments
'Big-Five' personality theory has reliably identified 5 main *traits* [15,16,21,26]: *Openness to Experience:* how curious/imaginative vs. cautious/habitual a person is, expressed by a preference for novelty and independence; *Conscientiousness*: organized vs. careless, expressed by self-discipline and goal seeking; *Extraversion:* outgoing vs. solitary*,* expressed by tendencies to seek out and enjoy others' company; *Agreeableness*: friendly vs. cold, expressed by co-cooperativeness and trust of others; *Neuroticism:* nervous vs. confident, expressed by how much one experiences negative emotions and is able to control these. Together these traits are known by the acronym OCEAN [21]. Standard reliable personality surveys have been developed where survey responses correlate with observable behaviors or trained analysts' evaluations for most traits [15, 26].

Classic personality theory involves inferring traits from *observable behaviors* such as language use, non-verbal communication, interview style, and appearance [15,26]. More recent personality research has broadened to examining *personal artifacts and physical spaces*, in part because traits such as Openness have proved more difficult to detect from observable behaviors.

Gosling [16,17] has systematically examined relations between personality and personal *physical* environments. In

each study, the owners ('targets') of that environment complete a standard personality survey. Next observers ('analysts') who do not know the owner inspect the target's environment, examining it for trait cues: for example how tidy it is, whether it contains photos of friends and family, diverse books and music, or inspirational posters. On the basis of their inspection, analysts fill out a personality survey assessing the target's personality. The key question is *whether analysts' personality assessments match the targets' self-evaluations*. For both dorms and offices analysts can identify Openness and Conscientiousness accurately. They are inaccurate at judging Agreeableness and moderately good at the remaining traits.

Gosling explains these relations between personality and physical environment in terms of three separate mediating mechanisms [16,17]. *Identity Claims* are when people *intentionally* structure personal environments to signal aspects of their personality to *others*. *Emotion Regulation* is *self-directed* organization occurring when people actively design personal environments to influence their mood. And *Behavioral Residue* is when people *unconsciously* leave informative traces in their environment following past actions. Behavioral Residue is a side-effect of everyday actions and is not *intentionally* created to affect self or others, in contrast to Identity Claims and Emotion Regulation.

Personal file systems are not generally intended for public viewing, or to be seen by others, so we anticipate that in our study, organization will be most affected by Behavioral Residue, somewhat affected by Emotion Regulation and relatively unaffected by Identity Claims.

Gosling's work also suggests possible *cues* for personality in the digital realm: Openness might be cued by a tendency to acquire more digital information, in particular varied types of files, e.g. music and pictures. Conscientiousness might be signaled by more systematic file organization. It is less obvious what might be reliable digital cues to Neuroticism, Agreeableness and Extraversion, and so our first study set out to explore this.

### Intentional Identity Claims: Profiles and Personality

Although there is little prior work on digital archives and personality, much recent work has addressed how people use other digital media, such as profiles to *intentionally* signal personal characteristics to others. People's Facebook and online dating profiles are actively crafted to signal personal characteristics [6, 11] and this has been explained using Goffman's account of self-presentation [14]. More indirect aspects of one's online profile such as number of friends, density of friendship networks or language used online also predict personality [1, 6, 18, 29]. Personality can also be detected from the contents of personal web pages [35]. Another approach has explored how personality relates to online preferences (e.g. Facebook 'likes') or the content people access. Personality can be inferred from browsing logs [9, 13, 20, 24]. Kosinski et al [24] analyzed

a massive collection of Facebook 'likes', finding that 'likes' predict various demographic, behavior and personality characteristics, most reliably Openness.

### Digital Collections

Some previous work has examined collections of *digital possessions* and shown that personality traits can be predicted from the content of music collections [31]. Other work explores how digital collections (particularly music and photos) relate to Identity and Emotion Regulation [27]. Finally, work on physical mementos examines how these are organized around the home to support different functions. Public mementos are placed in social locations and chosen to promote conversation or signal shared family values, whereas private ones are meant to engender deeper emotions and facilitate self-regulation [28]. However none of these studies of possessions has tried to relate the *structure* of digital collections to the owner's personality.

### STUDY 1: IDENTIFYING PERSONALITY CUES FROM STRUCTURE

This was an exploratory study to determine whether people ('analysts') could identify structural cues about a stranger's (target's) personality by examining their archive. Following Gosling [17], we expected analysts to be able to identify structural personality cues. We also wanted to see whether they could use these to accurately infer personality traits.
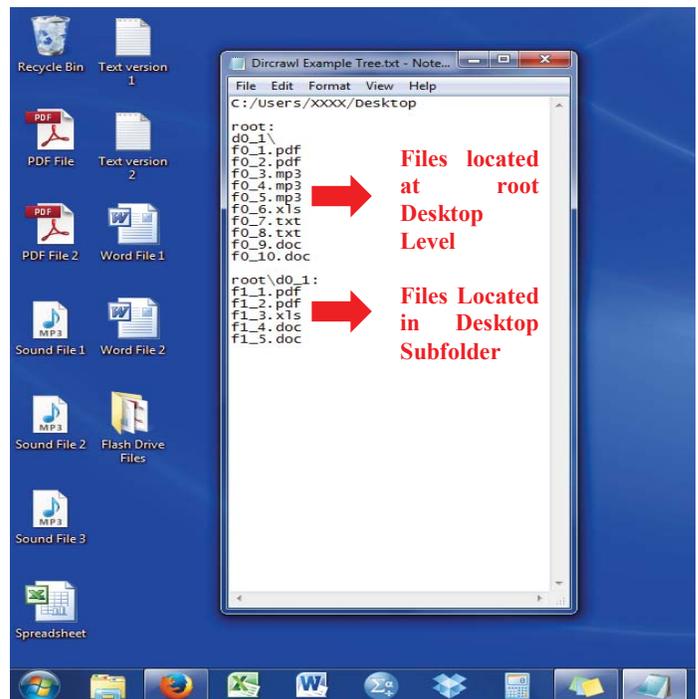
There were three steps to the procedure:



**Figure 1: Automatic structural analysis for the Desktop. Open window shows dircrawl automatic anonymized analysis of structure for desktop revealing: one folder and 10 objects that are unfiled on the desktop and one desktop subfolder containing 5 objects.**

- Analysts identified which structural attributes of the target's personal archive they thought served as *cues* to different personality traits,

- We determined targets' actual personality traits using a standard survey,

- Analysts judged the target's personality traits based on structural cues. We compared these with target's survey self-assessments allowing us to determine *judgment accuracy*.

## Method

*Profiling Targets and Generating Structural Information from Target Archives*

*Profiling Personality:* We first recruited targets and profiled their personality. We then programmatically analyzed their archives to generate cues for analysts. We administered the 44 item Big Five Inventory (BFI) [15, 21] to an initial 8 target volunteers, and from them we selected a subset of 4 targets who were distinct in terms of their Big 5 personality traits. This was to ensure variability across different traits in our targets, allowing us to assess multiple personality traits. Our procedure asked analysts to compare two target's traits and we wanted those traits to be distinct.

*Generating Structural Information:* For the selected 4 targets (who were all PC users), we extracted structural information about their digital archives by running a program ('dircrawl') that analyzes their file organization. The program also automatically anonymizes all file and folder names, while retaining file type information (e.g. docx, ppt, pdf). Retaining file types should allow an analyst to tell that a target has a large collection of movies or documents, but not the names or content of those files. Sample outputs for a target's desktop are shown in Fig. 1, along with the specific desktop that generated those outputs. The first line of the output in each cluster of files (beginning with "root:") shows the folder's location within the folder hierarchy followed by the anonymized names of the files in that folder. In this case the program was started at the desktop level, so the first cluster (under root:) is the desktop folder itself, which contains one folder (d0_1) and 10 files labeled (f0_1.pdf to f0_10.doc). Note that while each *name* is anonymized, file *type* (pdf, mp3, txt, doc etc.) is preserved. The next cluster (root\d0_1:) lists the contents of the subfolder the user created on the desktop (d0_1), which contains 5 objects f1_1.pdf to f1_5.doc.

We explained several sample program outputs to analysts, showing them how different file structures were represented in the program and walked them through a key, listing various file types. We did not provide summary outputs, e.g. total number of files, or mean files per folder, as we did not want to prejudice analysts about what might be significant structural cues.

To generate target data for the study, we used dircrawl to collect data from 4 separate folder locations for each target: Documents, Desktop, Pictures, and Videos folders. We also interviewed each target to rule out the possibility that they kept personal data in other idiosyncratic places. Information about music was not collected in this study because the majority of music files are automatically organized within applications such as iTunes, undermining the potentially predictive nature of personality and file/folder organization behaviors.

Analysts were only shown personal files. We excluded system files as these are not actively organized by the targets themselves [32]. Based on the different applications that our targets used, personal files were defined by inclusion as: Microsoft office files (e.g. doc, docx, xls, pptx, etc), Mac productivity files (e.g. pages, keynote, etc), picture files (e.g. jpg, png, tiff, etc.), music files (e.g. mp3, m4a, wav, etc.), video files (e.g. mov, wmv, avi, etc.), and zip, txt, mid, csv, rtf, wpd, msg, pdf, and mp4 files. We did not observe any targets using Open Office files (.odt, etc). Folders that contained only system files were also removed so as to avoid the appearance of spurious empty folders.

*Personality Trait Analysis*

*Participants:* Twelve analysts were asked to examine the structural information obtained by the dircrawl program and to use it to determine the target's personality type, giving their assessment of their Big 5 traits [21]. Analysts were mostly aged 18-21, drawn from a first year psychology class recruited via online sign-ups. They completed the study for course credit.

*Training on Structural Attributes:* We first explained the program outputs to the analysts, showing them various illustrative examples of non-anonymized personal archives alongside their structural output. We showed how the outputs provided information about file and folder structure as well as file types. Once analysts confirmed that they understood how outputs related to the original archives, we explained that their task was to use similar anonymized outputs to determine the target's personality traits.

*Training on Personality Traits:* We briefly explained the Big Five personality traits providing capsule summaries of each and ensured that analysts understood each trait's definition. Analysts were allowed to consult these personality trait definitions throughout the study and encouraged to ask clarifying questions at any point if they were unsure about either personality definitions or dircrawl program outputs.

*Procedure and Motivation:* Each analyst judged the personality of two targets. To provide motivation we told analysts to adopt the role of a Crime Scene Investigator recruited to provide expert character analysis derived from structural analysis of the digital archives. Analysts looked at the structural information on a computer. There were 4 separate dircrawls for each target (starting at Documents, Desktop, Pictures, and Videos folders), meaning that analysts looked at 8 different text files, presented two at a

time side-by-side (Desktops for the two targets, then Documents, etc.).

*Identifying Trait Cues:* To determine what analysts perceived to be important *cues* to personality, we elicited qualitative descriptions about the exact structural information were using to infer each personality trait. We also asked participants to compare and contrast the two targets for a given trait and explain how they used each target's structural information to formulate an assessment for that trait. These responses were audio recorded, and transcribed. Participants were allocated an hour for the task, although the average time taken was only 27 minutes.

*Trait Judgments:* Analysts also recorded their personality judgments on a set of 5 scales (one for each personality trait) where the ends of each scale represented the relevant trait anchors, e.g. Introvert-Extravert. Participants manually marked a point on a continuous scale for each target trait (e.g. if the analyst judged the target as an extreme Introvert they would give a score of exactly '1' and extreme Extravert a score of exactly '5' for that trait). We measured exact scale position for each trait, accurate to one sixteenth of an inch.

## Results

### Cues to Different Traits

We collected user comments about which aspects of structure signaled *cues* for each trait. Analysts noted multiple structural attributes as cues for a given trait and the same attribute was proposed for multiple different traits. The following content analysis was based on the 12 interviews. Consistent with Gosling [17] they provided more examples and were more confident about Openness and Conscientiousness than other personality traits.

Analysts readily provided cues for the *Openness* trait, and had clear views about how structure related to Openness. 11 analysts provided Openness cues, although one was uncertain about the reliability of the cue they provided. *Archive size*, particularly *overall number of pictures* and *videos* were all thought to signal Openness. Targets with more total files were thought to be more Open (6 analysts). Targets with many picture and video files were thought to be especially Open (8 analysts):

*A7: I rate that person a 5 (highest score) in Openness, just because they had **a lot of images and had a lot of files.***

Another important cue to Openness was general *lack of organization*; 4 analysts judged the target to be Open when their information was piled on the desktop or at shallow levels of the file hierarchy.

*A10: But Y is more Open, he has way more pictures, **even though they're not organized.***

Analysts were also able to identify clear cues signaling *Conscientiousness,* consistently focusing on *organization*. Altogether, 9 analysts provided cues to Conscientiousness.

All were confident about the cues they provided. Four described *general levels of organization*:

*A6 Their folders were **pretty organized**, so they were both pretty Conscientious.*

Three identified *greater use of folders*. The following analyst identifies a lack of Conscientiousness indicated by empty folders and unfiled items:

*A12: For Conscientiousness, **Y, had lots of empty folders and then there were tons of unsorted things.***

Unfiled desktop information indicated lower Conscientiousness:

*A11: X was more organized. And Y wasn't like a mess, but **they had a lot of stuff on their desktop***

Another analyst focused on *within* folder organization analyzing whether folders contained *consistent file types*.

*A1: X seemed pretty good at sorting things, so I guess that's Conscientiousness because a lot of their **folders had very similar file types.***

Analysts found it harder to identify cues to *Extraversion*. Nine provided cues, although two were unsure of what they proposed and one analyst failed to provide any cues, saying that identifying cues was too hard. As with Openness, 5 analysts felt Extraverts were more likely to have *more files* overall and, specifically, *more pictures*:

*A8: **Y has a lot of pictures and he has more documents**.*

Three analysts thought that Introverts were more likely to be *highly organized*, and Extraverts more haphazard:

*A12: I put by a 1 (extreme Introvert) based on the file structure **- it seems like they are less chaotic and less spontaneous and more organized.***

Analysts also found *Agreeableness* cues much harder to determine. Only 6 provided cues, with 3 of these being unsure, and 2 others saying it was hard to specify cues. There was a partial focus on *number of files* and *organization*, but often there was a lack of confidence in this judgment:

*A2: For Agreeableness, X seems like **they don't have many files** so, like, they don't want to be in the limelight I guess.*

In the same way, analysts found it hard to provide cues for *Neuroticism*. Only 6 provided cues, one of whom was uncertain about the cues provided. Three analysts were unable to provide any cues. Cues focused on *number of files* (4 analysts), and *level of organization* taken to indicate a desire to control and avoid anxiety.

*A12: X, maybe it could be like **they're super organized because they have a lot of anxiety, which causes them to be orderly**.*

### Analysts Trait Judgments Are Accurate

*Comparison between Analyst and Target Evaluations:* While the focus of this study was to generate cues we also explored the *accuracy* of analysts' judgments based on

these cues. For all 12 analysts, we correlated *analyst's evaluations* of each trait with the ground truth of the *target's own self-evaluation* of that trait. Despite having just 24 observations (2 for each of 12 analysts), correlations were significant for Openness $r(22)=0.412$, $p=0.045$. None of the other individual trait correlations were significant. However when we combined *all traits* in a single correlation, judgments were accurate. After removing missing data points, for all 5 traits combined, there was a significant correlation: $(r(112)=0.224, p=0.016)$.

Overall the results are encouraging, suggesting analysts can both identify structural cues and use these to accurately infer personality traits particularly Openness. Although Conscientiousness judgments were inaccurate participants were nevertheless able to generate plausible cues for this trait. Analysts were unconfident about cues for other traits, and their judgments less accurate.

## OBJECTIVE RELATIONS BETWEEN PERSONALITY AND PERSONAL ARCHIVE STRUCTURE

The first study suggested analysts could identify cues and infer some personality traits from structural information about a stranger's archive. Our second study explored this relationship more directly by assessing whether specific *self-reported personality traits* predicted *objective structural characteristics* of a person's personal archive.

Analysts in our first study oriented to the following cues: (a) total number of files, (b) extent to which people imposed organization on these files by foldering them, (c) use of privileged locations such as the Desktop, and (d) number of media files such as music or photos. Results suggested that these were reliable cues to Openness, so we evaluated these in study 2. We explored Conscientiousness as our analysts' comments and prior work [17] suggest this trait might be inferable from archival properties. In addition we explored a potential cue for Neuroticism based on a trend observed in previous pilots of this study.

### Method

*Participants and Personality Assessment:* We collected personality profiles from 62 participants (23 PC users and 39 Mac users). We obtained self-reported personality traits using the same Big Five Personality survey as in the first study. Again participants were recruited from a first year psychology class and their ages were 18-21. All participants had owned their laptop for at least one year.

*Analyzing Personal Archives:* We used the same dircrawl program as the first study, run on four major file locations (Documents, Desktop, Pictures, and Music). In contrast to the first study, we analyzed Music rather than Videos as both the literature and the participants in our first study suggested Music was important, and few of our participants kept large numbers of videos. To obtain informed consent from our participants, we first demonstrated our dircrawl analysis program showing how it generates anonymized structural data. Participants were encouraged to ask questions and allowed to withdraw if they did not want to participate. None withdrew.

Participants were asked about other locations where they actively organized personal files. One participant mentioned Downloads, but we excluded this because they did not actively organize information within it. Another mentioned Dropbox, but we this was excluded because as a shared resource it may have been organized by others. Files in My Music and My Photos folders are an ambiguous case; they are personal files, but are often automatically organized by the system, such as iTunes or iPhoto. Files from these locations were included in measures of total number of files, but excluded from analyses of active organization due to the higher likelihood that they had been automatically organized for the user.

*Properties of Archives:* We examined the following structural variables derived from dircrawl:

*Total files*: All personal files located in all four default folder locations combined (Desktop, Documents, Music, and Pictures). This gave an overall measure of the *size* of the person's archive.

*Total files in actively organized locations:* All files in the Documents and Desktop folders. We are interested in active organization so this variable excluded files from the Pictures and Music folders as these were more likely to be automatically organized (e.g. iTunes, iPhoto, etc).

*Files/folder:* Average number of files per folder in actively organized locations (Documents and Desktop). This gave us a measure of a person's propensity to impose organization on actively managed information.

The Desktop's salience makes it a privileged location for personal archives. It is actively used for task management and reminding about outstanding tasks [25, 37]. We therefore included two specific measures of Desktop usage.

*Total desktop files:* Number of files located in the entire Desktop folder.

*Percentage of Desktop files in organized folders:* Percentage of total desktop files that are located in subfolders rather than scattered unorganized at the root level of the Desktop space. This provides a measure of the extent to which the desktop is organized versus cluttered.

Gosling's work [17] also suggests the importance of media such as music and books in revealing personality traits, so we also measured:

*Total music files:* Total number of music files across all folder locations.

*Total picture files*: Total number of picture files across all folder locations.

All of these archival metrics were highly skewed, so, to run analyses to determine their relationships with personality traits, they were each log transformed.

| Archive Property | Mean | Median | Std. Dev. |
|---|---|---|---|
| Total Files | 15320.1 | 9330 | 24174.2 |
| Total Files in Actively Organized Locations | 5395.9 | 864 | 21925.5 |
| Files/Folder in Actively Organized Locations | 18.9 | 13.2 | 16.4 |
| Total Desktop Files | 4745.2 | 272 | 21827.9 |
| Percentage Desktop Files Organized in Folders | 75.1 | 98.3 | 40.5 |
| Total Picture Files | 11287.3 | 5239 | 19214.6 |
| Total Music Files | 2929.45 | 1551 | 3994.5 |

**Table 1: Descriptive statistics prior to log transformation**

Our predictions are derived from our first study and Gosling's findings about personality in physical environments [17]. Based on Study 1 and Gosling's findings we expected Openness and Conscientiousness to predict PIM practices. We were also interested in looking into a potential relationship between Neuroticism and the desktop based on a trend seen in prior pilots of this study. We repeatedly observed that Neurotic participants kept more items on the Desktop possibly to reduce anxiety about forgetting ongoing tasks involving those items. Predictions and results are presented separately for each trait.

### Results

Descriptive statistics prior to log transformation of each variable are presented in Table 1.

*Conscientiousness Predicts Organization*

Conscientious people have clean, organized, neat, and uncluttered physical spaces [17]. Our initial study also suggested Conscientiousness cues such as having "more things organized," as well as "[having] less stuff." Lack of conscientiousness was signaled by "a lot of files on their desktop." We therefore predicted Conscientious individuals would have:

- *Fewer files overall*, because they don't keep superfluous information.
- *Fewer files per folder* because they should be more actively organizing.
- *Fewer files* on the *Desktop,* to keep their workspace clear.
- *Higher percentage of desktop files organized into folders* to actively manage ongoing tasks.

We used regressions to test whether total number of files in actively organized locations, number of files per folder, number of desktop files, and percentage of foldered desktop files are significantly related to Conscientiousness. Variables were also centered to control for multi-collinearity. We conducted one regression rather than multiple correlations to avoid the risk of generating spurious significant results following multiple statistical tests.

As OS affects organization [5], we conducted separate analyses for PC and Mac users. We tested 2 models for each. Model 1 tested Total Files and Files per Folder as predictors. In addition we tested the interaction, Total Files X Files per Folder, to determine whether tendency to organize is mediated by total number of files. Model 2 adds desktop variables: Total Desktop files and Percentage of foldered Desktop files.

For PC users (see Table 2) Model 1 is not significant ($R^2$=.108,F=0.689,p=.571), but if we include Desktop variables the regression is highly significant ($R^2$=.709,F=7.297,p=.001), and all variables have significant effects. Model 2 indicates Conscientious PC users keep fewer files overall, confirming our first hypothesis. However, files per folder is *positively* related to conscientiousness, which contradicts what we expected. One explanation is that Conscientious individuals are less likely to create 'failed folders' [36] that are empty or contain one or two files only. Again contradicting our hypothesis, Conscientious people keep *more* files on the Desktop, possibly to better manage their tasks. However as expected they are active in foldering these. Fig 3 shows the

| | Model 1 General Organization | | | Model 2 Desktop Organization | | |
|---|---|---|---|---|---|---|
| | Coef. | SE | P | Coef. | SE | P |
| Intercept | 3.611 | .152 | .000 | 2.280 | .262 | .000 |
| Total Files (Documents & Desktop) | .012 | .083 | .884 | -.362 | .087 | .001 |
| Files/Folder | .243 | .200 | .240 | .443 | .128 | .004 |
| Total Files X Files/Folder | .110 | .096 | .270 | -.148 | .060 | .026 |
| Desktop Files | | | | .145 | .053 | .015 |
| Desktop Files in Self-Created Folders | | | | .212 | .062 | .004 |
| | $R^2$ | F | | $R^2$ | F | P |
| | .108 | .689 | .571 | .709 | 7.297 | .001 |

**Table 2: Nested lagged regression of the general organization metrics and the desktop organization metrics related to Conscientiousness for PC users.**

interaction between total files and active organization indicating that the relationship is moderated by the total number of files, with files/folder only being predictive of Conscientiousness when people have low numbers of files. This may be explained by the fact that it is more cognitively demanding to systematically folder when one has large numbers of files, making it hard even for Conscientious participants to keep data organized.
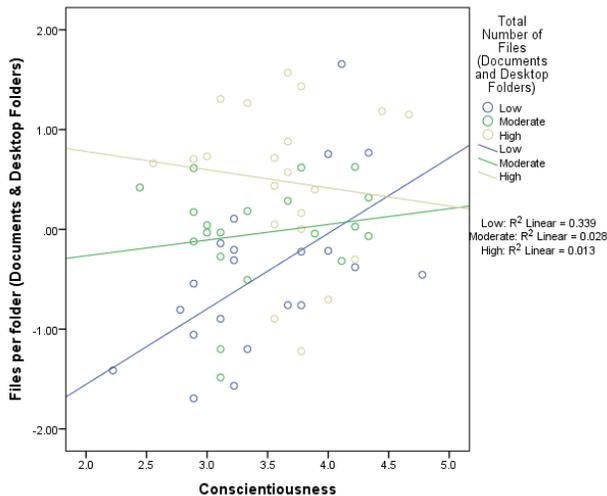


**Figure 3: Relations between Total Number of files and Files per folder and Conscientiousness: Conscientiousness predicts active foldering only when people have fewer files**

For Mac users the regression analysis was only significant for Model 1 ($R^2$=.222,F=3.327,p=.031). The interaction was the only significant predictor, indicating a similar mediating relationship between total number of files and files per folder as was found for PC users. However, including the desktop terms did not improve the regression. This suggests that Mac users have a very different relationship with the desktop space. Whereas for PC users it is a space that seems to be related to Conscientiousness, for Mac users this is not the case.

*Openness*
Cues for identifying Openness in physical spaces include distinctive, stylish, and unconventional spaces and varied CDs and magazines [17]. Participants from our initial study also reported that they relied on the total number of files for determining Openness, although this is not what Gosling found. From this we hypothesize that:

- **Open participants should have more *total files, more music and more picture files* because of a desire to expose themselves to a broad range of stimuli**

- **Open individuals should have more varied *types* of files**

Regression analyses were run, again separating Mac and PC users, to test whether total number of files overall, total number of picture files, and total number of music files are significantly related to Openness. We again include an interaction term to allow for the fact that increased number of music and picture files may mean different things based on the total number of files overall. Variables were again centered to control for multi-collinearity.

Contrary to our predictions, neither model was significant, For PC users: ($R^2$=.350,F=1.829,p=.161) and for MAC users ($R^2$=.120,F=.896,p=.495) confirming Gosling's findings that Openness is not predicted by total files, amount of music, and amount of pictures. Gosling's findings would argue instead that it is the *variety* of music and pictures that indicate Openness. However filetype variety was unrelated to Openness ($R^2$=.031,F=.568,p=.572) for MACs and ($R^2$=.165,F=1.876,p=.181) for PCs.

*Neuroticism*
Neuroticism is not well predicted in physical spaces [17]. Participants from our initial study also reported difficulties identifying cues to Neuroticism. However, other pilot studies we have conducted consistently suggest a relationship between neuroticism and use of the desktop space. This may be because anxious people keep more of their files on the desktop to avoid forgetting active to-dos. Based on this our prediction for neuroticism is:

- **Neurotic participants should have more desktop files because of concerns about forgetting important tasks.**

We conducted a simple regression combining all users to see if total number of desktop files was a significantly related to Neuroticism. The regression was not significant ($R^2$=.043,F=2.706,p=.105) but it does appear to be trending in the predicted direction. However one characteristic of the desktop is that its 'messiness' may fluctuate with workload. Given that our sample was made up of a student population we took into account students' work cycle where the beginning of an academic term marks a reset. We expected desktop effects to be most pronounced for students tested late in the quarter as work piles up. When we ensured that we sampled participants tested at least halfway through the academic term (N = 45) this same regression becomes significant ($R^2$=.144,F=7.418,p=.009) indicating that high number of files on the desktop relates to Neuroticism. Future iterations of this work will be necessary to confirm the validity of this relationship.

**DISCUSSION AND CONCLUSIONS**
Our finding that people's personality affects how they organize is important for PIM theory. It demonstrates that PIM behaviors are not entirely *functional*. Rather than someone's file system being rationally organized to maximize retrieval, our results suggest that organization is partially driven by aspects of the users' personality. However it could still be true that a match between personality and organization may optimize retrieval and future work could test this.

This study presents a new approach to understanding individual differences in PIM practices. There are various technical implications to these findings. Personality factors may help address pernicious problems in the *design* of PIM tools, allowing us to target tools to different user traits. A well-documented barrier to the widespread adoption of PIM tools is that users employ very different strategies for the same organizational tasks [23,36]. We have shown that personality is one source of these differences, although other factors such as job type [3] may also be important. Our results allow us to predict organizational style once we have profiled personality. Tools might therefore provide personality based defaults based on different user personality traits. Thus, it may be that for PCs, Conscientious users may want tools that provide greater support for task management and the Desktop. Neurotics may also want to organize ready access to their Desktops. While personalization has been proposed previously, as far as we are aware, this is the first research to make *specific recommendations* about how it might be implemented to respect personality differences. There is also growing concern about how personal information management practices can cause knowledge workers stress and anxiety [8]. Devising PIM practices that suit one's personality might also reduce such stress.

We also contribute to new theoretical approaches to the study of personality that examine *possessions* rather than *observable behaviors* as signals of personality. While prior personality work has explored mechanisms of Identity Claims and Emotion Regulation [1, 16, 17, 29], we believe our study is the first to look at digital *Behavioral Residue*. File systems are usually private, and we extend other work demonstrating relationships between traits and *public* aspects of digital behavior such as active construction of profiles, or active digital behaviors such as browsing, language use or Facebook 'likes' [1, 6, 18, 29]. Conscientiousness was related to smaller archives and more Desktop files with these being organized into folders, presumably to support reminding and task management.

We also found interesting differences between PC and MAC users; for PC users, systematic active use of the desktop related to Conscientiousness, while this was not true for MAC participants. Another intriguing aspect of our findings concerned the interaction between traits and the total amount of information stored. While Conscientious people stored less information overall, the effects of Conscientiousness are most marked for participants who stored lower amounts of information. These people seemed less likely to create spurious unused folders. We also found that as workload increased, Neurotics reacted by keeping more files on the desktop. This could be to decrease anxiety over forgetting pending work tasks and to-dos.

There are limitations to our study. Our analytic program Dircrawl does not provide *spatial* organizational information which some users thought could be related to personalitytraits. Future work should explore this. There is also evidence that people are increasingly storing materials in the *cloud* [32], which we didn't examine here. We also focused on actively organized archives, but it might also be interesting to include temporary locations such as Downloads or Trash, although there are clearly issues about when to record representative usage, as many participants periodically clear these out. In addition our participant pool was homogeneous and more diverse populations should be explored.

Another issue concerns intriguing inconsistencies between our two studies. Why were analysts able to accurately detect Openness in study 1, but we could not find relationships in study 2? It seems that our analysts had access to information that was not captured by the summary statistics tested in study 2. Prior work suggests *variety* of music and magazines are strong Openness cues, and it may be that our hypotheses need to be refined. We carried out a simple analysis of variety of file *types,* but it may be that other properties such as age, length of file name, creator and so forth might offer more reliable indicators of variety and hence Openness.

The opposite was true for Conscientiousness. We were able to detect relations between objective organization and Conscientiousness. Intriguingly, study 1 shows analysts clearly oriented to the correct structural properties although their final judgments were inaccurate. This may be because certain relations are unexpected. Our regression model showed counterintuitive relations, e.g. Conscientious people have *more* desktop files and *more* files per folder. Our findings on Neuroticism make intuitive sense although they are not predicted by theory, and need to be explored further.

An interesting future direction concerns the organization of *shared* tools such as Dropbox or GoogleDrive. Prior work suggests that collaborators experience difficulties in co-organizing such folders [30, 32, 33]. However theoretically we could examine whether shared tool organization reveals traces of *all* participants' personalities or whether more *dominant* team members' personality is most clearly stamped on the shared archive.

Future work should continue to explore these novel relations between personality and personal digital archives. That work could extend personality theory exploring relations between possessions and personality. Enhanced understanding will also address outstanding issues surrounding personalization and customization that has been a substantial barrier to the adoption of general PIM tools.

## REFERENCES

1. Bachrach, Y., Kohli, P., Graepel, T., Stillwell, D.J, Kosinski, M. Personality and patterns of Facebook usage. *Proc WebSci*, ACM Press (2012) 36–44.

2. Barreau, D., & Nardi, B. A. Finding and reminding: file organization from the desktop. *ACM SigChi Bulletin*, *27*(3) (1995) 39-43.

3.  Bellotti, V., Ducheneaut, N., Howard, M., Smith, I., & Grinter, R. E. Quality versus quantity: E-mail-centric task management and its relation with overload. *HCI*, *20*(1) (2005), 89-138.

4.  Bergman, O., Beyth-Marom, R., Nachmias, R., Gradovitch, N., & Whittaker, S. Improved search engines and navigation preference in PIM. *ACM (TOIS)*, *26*(4) (2008), 20-58.

5.  Bergman, O., Whittaker, S., Sanderson, M., Nachmias, R., & Ramamoorthy, A. The effect of folder structure on personal file navigation. *Journal of the ASIST*, *61*(12) (2010) 2426-2441.

6.  Boyd, D. Friendster and publicly articulated social networking. *Proc. CHI* (2004), 1279–1282.

7.  Cutrell, E., Robbins, D., Dumais, S., & Sarin, R. Fast, flexible filtering with phlat. *Proc CHI,* ACM Press. (2006) 261-270.

8.  Dabbish, L. A., & Kraut, R. E. Email overload at work: an analysis of factors associated with email strain. *Proc. CSCW,* ACM Press(2006) 431-440.

9.  De Bock K, Van Den Poel D (2010) Predicting website audience demographics for Web advertising targeting using multi-website clickstream data. *FI 98(1)* (2010) 49–70.

10. Dumais, S., Cutrell, E., Cadiz, J. J., Jancke, G., Sarin, R., & Robbins, D. C. Stuff I've seen: a system for personal information retrieval and re-use. *Proc. SIGIR,* ACM Press. (2003) 72-79.

11. Ellison, N., Heino, R., & Gibbs, J. Managing impressions online: Self-presentation processes in the online dating environment. *JCMC 11*, 2 (2006), Article 2.

12. Fisher, D., Brush, A. J., Gleave, E., & Smith, M. A. (2006, November). Revisiting Whittaker & Sidner's email overload ten years later. *Proc. CSCW*, ACM Press (2006) 309-312.

13. Goel S, Hofman JM, Sirer MI (2012) Who does what on the Web: Studying Web browsing behavior at scale. *Proc ICWSM* ACM Press (2012) 130–137.

14. Goffman, E. *The presentation of self in everyday life*. Anchor: New York, 1959.

15. Goldberg, L. R. The development of markers for the Big-Five factor structure. *Psychological assessment*, *4*(1) (1992) 26-39.

16. Gosling, S. *Snoop: What your stuff says about you*. Basic Books. (2009)

17. Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. A room with a cue: personality judgments based on offices and bedrooms. *Journal of personality and social psychology*, *82*(3) (2002)  379-394.

18. Gosling, S. D., Gaddis, S., & Vazire, S. Personality Impressions Based on Facebook Profiles. *Proc. ICWSM* ACM Press (2007) 123-130.

19. Gwizdka, J. Email task management styles: the cleaners and the keepers. *Proc. CHI,* ACM Press. (2004) 1235-1238.

20. Hu J, Zeng H-J, Li H, Niu C, Chen Z. Demographic prediction based on user's browsing behavior.  *IWWWC,* (2007) 151–160.

21. John, O. P., Donahue, E. M., & Kentle, R. 'The Big Five: Factor Taxonomy. *In Handbook of Personality: Theory and Research* (1990) 66-100.

22. Joinson, A.N. Looking at, looking up or keeping up with people?: Motives and use of Facebook. *Proc. CHI* ACM Press (2008) 1027–1036.

23. Jones, W. *Keeping Found Things Found*. Morgan Kaufmann. (2010).

24. Kosinski M, Kohli P, Stillwell DJ, Bachrach Y, Graepel T. Personality and website choice. *ACM WebSci* (2012) 251–254.

25. Malone, T. W. How do people organize their desks?: Implications for the design of office information systems. *ACM TOIS*, *1*(1) (1983) 99-112.

26. McCrae, R. R., & Costa, P. T. Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology*, *52*(1) (1987) 81.

27. Odom, W., Pierce, J., Stolterman, E., & Blevis, E. Understanding why we preserve some things and discard others in the context of interaction design. *Proc .CHI*, ACM Press. (2009) 1053-1062).

28. Petrelli, D., Whittaker, S., & Brockmeier, J. AutoTopography: what can physical mementos tell us about digital memories? *Proc CHI*, ACM Press (2008) 53-62.

29. Quercia D, Lambiotte R, Kosinski M, Stillwell D, Crowcroft J (2012) The Personality of popular Facebook users. *Proc CSCW,* ACM Press (2012) 955–964.

30. Rader, E. Yours, mine and (not) ours: social influences on group information repositories. *Proc CHI,* ACM Press (2009) 2095-2098.

31. Rentfrow, P. J., & Gosling, S. D. The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of personality and social psychology*, *84*(6), (2003) 1236-1258.

32. Tang, J., Drews, C., Smith, M., Wu, F., Sue, A., & Lau, T. Exploring Patterns of Social Commonality Among File Directories at Work. *Proc CHI,* ACM Press (2007) 951-960.

33. Tang, J., Brubaker, J., & Marshall, C. What Do You See In The Cloud? Understanding the Cloud-Based User Experience through Practices. *HCI – INTERACT* (2013) 563-572.

34. Teevan, J., Alvarado, C., Ackerman, M. S., & Karger, D. R. The perfect search engine is not enough: a study of orienteering behavior in directed search. *Proc CHI,* ACM Press (2004) 415-422.

35. Vazire, S., & Gosling, S. D. e-Perceptions: personality impressions based on personal websites. *Journal of personality and social psychology*, *87*(1), (2004) 123-132.

36. Whittaker, S., & Sidner, C. Email overload: exploring personal information management of email. *Proc. CHI,* ACM Press (1996) 276-283.

37. Whittaker, S.  Personal information management: from information consumption to curation. *ARIST*, *45*(1) (2011) 1-62.